NANOSENSORS
PHYSICAL, CHEMICAL, AND BIOLOGICAL

Series in Sensors

Series Editors: Barry Jones and William B. Spillman

Other recent books in the series:

NANOSENSORS
PHYSICAL, CHEMICAL, AND BIOLOGICAL

Vinod Kumar Khanna

CSIR-CEERI, Pilani, India

Dedicated to the loving memory of my late father

Shri Amarnath Khanna

for nurturing my educational career.

Also dedicated to my mother Shrimati Pushpa Khanna,

my daughter Aloka Khanna, and my wife Amita Khanna

for their affection and support.

Contents

Preface

Objectives and Goals

Interesting developments have taken place in the field of sensors in recent years. Nanotechnology holds a great potential for revolutionizing the sensor arena. Traditional sensors are being reengineered and recast from a nanotechnological perspective, and new sensor designs are being introduced. Sensitivity, detection range, and response times of sensors have shown remarkable improvement by using nanotechnological methods. Faster, better, cheaper, and smaller sensors are becoming available.

Sensor development in the nanotechnology age, the focal theme of this book, is one of the hottest topics in contemporary research. The book presents a critical appraisal of the new opportunities provided by nanotechnologies and the nanotechnology-enabled advancements in the field of sensors.

Rationale

Being a new topic, information on nanosensors is scattered across research journals and reports. There is a paucity of books in this area. This book intends to fill this long-felt need. Bringing together the widely spread information between the covers of an easy-to-read book will be helpful to all those working on nanosensors for instant reference and updates.

Topical Coverage

Chapter 1 provides an introductory survey of nanotechnology and introduces preliminary background material for understanding the rest of the book. It thus lays the groundwork for the book.

Chapter 2 describes the materials utilized for the fabrication of nanosensors such as nanoparticles made of metals, semiconductors and insulators, quantum dots, and carbon nanotubes. These constructional nanomaterials constitute the building blocks of nanosensors.

In Chapter 3, the vision of a laboratory working on nanosensor fabrication and characterization is sketched. Such a laboratory has a broad scope

and pools resources from vastly different areas. Integration of these facilities under one roof, though, is an uphill task.

The remaining chapters address various categories of nanosensors. Chapter 4 is dedicated to mechanical nanosensors, e.g., MEMS and NEMS resonators, single-electron transistor-based displacement sensors, electron tunneling accelerometers, CNT force, pressure and flow sensors, etc.

Chapter 5 deals with temperature sensors such as miniaturized thermo-couples, CNT resistors, ratiometric fluorescence intensity sensors, hot electron bolometers, and so on.

Chapter 6 provides a detailed description of optical nanosensors like those based on surface plasmon resonance, SERS scattering, colorimetric sensors, fiber-optic nanosensors, PEBBLEs, QD devices with special emphasis on FERET-type sensors, etc. Optical nanobiosensors are also covered in this chapter.

Magnetic nanosensors, the subject of Chapter 7, have received a boost with the discovery of the giant magnetoresistance effect. Many landmark nanobiosensors have been developed and some of these are discussed in this chapter. Magnetic relaxation is another phenomenon that has been widely applied in designing nanobiosensors. This is briefly dealt with here.

Chapter 8 describes the coveted "nanobiosensors," which unify the living world with physical and chemical domains. The "marriage" of biomolecules and nanotechnology often yields novel sensor concepts. Such biosensors have a distinct impact in different fields of application such as medicine, food technology, environment, chemistry, and biotechnology, as well as information processing. These nanosensors, as well as all others, have been the subject of intensive research globally in recent years. There has been a veritable deluge of research publications in this vital area because it is directly related to human health and well-being. Electrochemical nanobiosensors are comprehensively surveyed. Many kinds of nanoparticle-modified electrodes are also reported. Apart from these nanosensors, mechanical nanosensors such as those based on micro- or nanocantilevers are outlined, along with additional optical nanosensors that use aptamers.

Chapter 9 deals with chemical nanosensors primarily aimed at detecting gaseous compounds, elements, and parameters such as pH of aqueous solutions. Metallic nanoparticles, metal oxide nanoparticles, and CNTs, pristine or functionalized, have been used to develop a family of sensors for chemical analysis, which will be useful for analytical chemists. *Chemiresistors* fabricated on MEMS hotplate with nanostructured tin oxide film coating provide high sensitivity for gas detection. Progress in nanomaterials has created tremendous opportunities for improved gas sensor devices. In nanoparticle-based semiconductor gas sensors, the detection thresholds for gases have been lowered severalfold compared to currently commercialized sensors.

The concluding chapter, Chapter 10, reviews and introspects on the main developments in the different varieties of nanosensors that have been

described in this book. The present scenario is assessed, and, on the basis of this panoramic view, the main research areas calling attention are highlighted. Core problems facing researchers, notably the interfacing of nanosensors with the macro world and related issues, are discussed.

Style and Presentation

Nanosensors, like their conventional macro- and micro-counterparts, are an interdisciplinary field that attracts scientists, engineers, and students alike. To cater to this wide audience, a thoroughly grassroot approach has been adopted in this book. Professionals in one area are not expected to be conversant with the terminology of another discipline, and, therefore, the responsibility to explain the key terms of different areas has been clearly borne in mind. It is earnestly hoped that the subject matter will be understandable to personnel specializing in different fields. Another noteworthy feature of this book is that it is not written in a dull, monotonous vein but in a question–answer format to keep the reader's interest alive. Thought-provoking questions are raised time and again to arouse the reader's interest.

Throughout the book, end-of-chapter exercises are appended to rekindle the reader's interest. Further, worked examples are sprinkled throughout the book to supplement the text.

Intended Audience

It is anticipated that the book will provide a useful forum for a wide cross section of readers, including scientists, engineers, teachers, and students. It will serve as a useful resource, both for research and teaching purposes.

Vinod Kumar Khanna
CSIR-CEERI, Pilani, Rajasthan, India

Acknowledgments

Author

Vinod Kumar Khanna received his MSc from the University of Lucknow in 1975 and his PhD from Kurukshetra University in 1988 for the thesis entitled "Development, Characterization and Modeling of the Porous Alumina Humidity Sensor." He joined the Solid-State Devices Division of CSIR-CEERI on April 21, 1980. He has been working for the last 30 years on the design, fabrication, and characterization of various solid-state devices, mainly power semiconductor devices and sensors. He has been actively involved in different research and development projects on thin film humidity sensor, high-voltage TV deflection transistor, power Darlington transistor for AC motor drives, fast-switching thyristor, high-current and high-voltage rectifier, neutron dosimetry diode, power DMOSFET and IGBT, PMOSFET gamma ray dosimeter, MEMS technology–based microsensors, ion-sensitive field-effect transistor (ISFET), ISFET-based chemical and biosensors, capacitive MEMS ultrasonic transducer (cMUT), acoustic sensor, pressure sensor, microcantilever, MEMS gyroscope, and MEMS hotplate gas sensors.

Dr. Khanna's deputations abroad include Denver, Colorado, USA, for presentation of research papers in *IEEE-Industry Applications Society Annual Meeting*, 1986; Technical University, Darmstadt, Germany, 1999, under CSIR-DLR collaborative project; Kurt-Schwabe-Institut für Mess- und Sensortechnik e.V. Meinsberg, Germany, 2008, under CSIR-DAAD Exchange of Scientists Programme; and Institute of Inorganic Chemistry, Novosibirsk, Russia, as a member of the Indian delegation in an Indo-Russian workshop.

He has authored six books of which *IGBT Theory and Design* has been published by IEEE Press-Wiley Interscience, New Jersey, 2003. He has also published 115 research papers in various reputed international/national journals and conference proceedings. He holds three patents.

Dr. Khanna is a fellow of the Institution of Electronics and Telecommunication Engineers, India. He is a life member of the Indian Physics Association (IPA); presently, he is chairman of IPA, Pilani Chapter. He is also a life member of the Semiconductor Society (India) and the Indo-French Technical Association. Citations include *Marquis Who's Who in Science and Engineering*, *Marquis Who's Who in Asia*, etc.

His fields of specialization are power semiconductor devices, MEMS, and microsensors. Presently, he is Chief Scientist and Head of MEMS and Microsensors Group at CSIR-CEERI.

1

Introduction to Nanosensors

1.1 Getting Started with Nanosensors

This book deals with *Nanosensors: Physical, Chemical, and Biological.* So, the questions that immediately come to mind are as follows: *(1) What is nanotechnology? (2) What are nanosensors? (3) What is the meaning of the terms "physical," "chemical," or "biological" alone and when applied to nanosensors?*

These questions need to be answered at the outset. Clearly, the starting point in the study of nanosensors will be the basic concepts of physics, chemistry, biology, and nanotechnology. Semiconductor electronics will also be introduced. The key terms will be defined. These will enable the reader to recapitulate the preliminary knowledge required to understand the subject matter. They will also help to build up the vocabulary of the new words of nanotechnology that the reader will come across in this book. An exhaustive revision of these fields is not intended, only reminisces of interesting physical, chemical, and biological materials and phenomena will be provided to highlight the interdisciplinary nature of the subject.

Then the motivation behind advancement from macro- to micro- and to nanosensors will be explained. The reader will be introduced to the rapidly advancing field of nanosensors and acquainted with different types of nanosensors, treating the fundamental principles and classification of nanosensors and laying the foundation upon which the succeeding chapters will be constructed. The scope and organizational plan of the book will be described.

Let us begin by revising the knowledge of elementary natural sciences.

1.2 Natural Sciences

What is natural science? Natural science is the systematized knowledge of nature and the physical world. It is a science, such as biology, chemistry, physics, and earth science, that deals with objects, phenomena, or laws of nature and the physical world (Brooks 2006).

1.3 Physics

1.3.1 Definition of Physics

Physics is the science of forces that exist between objects, the interrelationship between matter and energy, and interactions between the two (Daintith 2010).

1.3.2 Branches of Physics

What are the different branches of physics? Physics is grouped into traditional fields such as mechanics, properties of matter, heat, light, sound, electricity, and magnetism, as well as in modern extensions including atomic and nuclear physics, cryogenics, solid-state physics, particle physics, and plasma physics.

1.3.3 Matter: Its States, Materials, and Particles

Matter is something that has weight and occupies space, for example, air, water, gold, iron, and wood. It exists as a *solid* having a definite volume and shape, for example, a block of copper; a *liquid* having a definite volume but not a definite shape, such as milk and juice; a *gas* having neither a definite volume nor a definite shape, for example, nitrogen and oxygen; or a *plasma* containing a mixture of positive ions and free electrons in almost equal proportions.

   Materials are substances out of which things are or can be made, serving as inputs to manufacturing plants. A *particle* is a very small piece or part, a tiny portion or speck of a material.

1.3.4 Molecules, Atoms, and Atomic Structure

A *molecule* is the smallest particle of a substance, element (which cannot be broken down into simpler substances), or compound (made of two or more simpler substances) that retains the chemical properties of the substance. An *atom* is the smallest part or particle of an element, having all the chemical characteristics of that element that uniquely define it, and can take part in a chemical reaction.

   Does an isolated atom or molecule possess the bulk physical properties of the element or compound from which it is made? Not necessarily, because the influence of neighboring atoms or molecules, as present in bulk form, is no longer exercised.

   Is an atom the ultimate smallest particle or are there further smaller particles? An atom is not the ultimate smallest particle into which matter is broken down. Still smaller particles exist. Then, *what is the structure of the atom?* Talking about atomic structure, an atom consists of a central nucleus surrounded by a cloud of one or more electrons (Figure 1.1). The electron has a rest mass
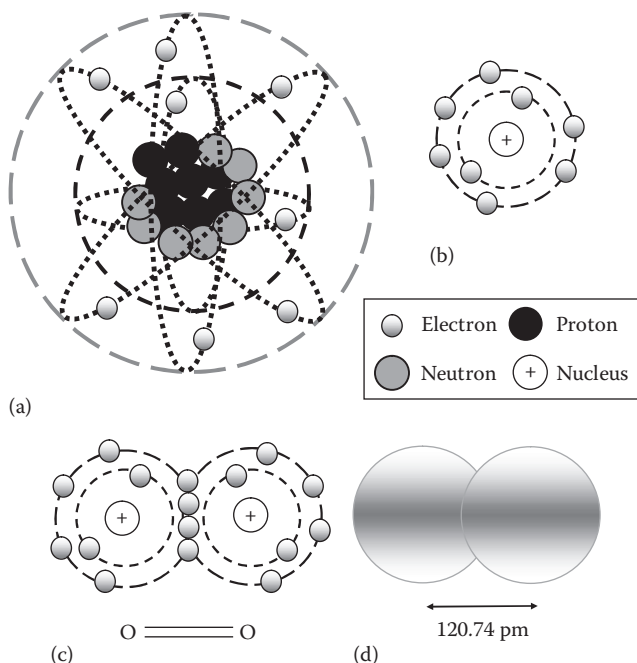
FIGURE 1.1
(a) Structure of oxygen atom (atomic number 8, atomic weight 16) containing 8 protons ($1s^2\,2p^6$) in the planetary atomic model. (b) Shell model of oxygen atom with 2 electrons in shell 1 and 6 electrons in shell 2. (c) Line diagram of oxygen molecule showing double bonds each consisting of a pair of shared electrons. (d) Two orbitals in oxygen molecule. An orbital is the region surrounding an atomic nucleus indicating the electron's probable location.

m_e of 9.1066×10^{-28} g and a unit negative electric charge of 1.602×10^{-19} C. The *nucleus* is positively charged and contains one or more relatively heavy particles known as *protons* and *neutrons*. A proton is positively charged. The number of protons in the nucleus of an atom is the *atomic number* for the chemical element. A proton has a rest mass, denoted by m_p, of approximately 1.673×10^{-27} kilogram (kg). A neutron is electrically neutral and has a rest mass, denoted by m_n, of approximately 1.675×10^{-27} kg.

The number of protons in the nucleus of an atom is its atomic number (Z). The total number of protons and neutrons in the nucleus of an atom gives its *mass number*, *atomic weight* (A), or *relative atomic mass*. It is the ratio of average mass per atom of the naturally occurring form of an element to 1/12th the mass of a carbon-12 isotope; an *isotope* is one or more atoms of an element having the same atomic number but different atomic weight. One-twelfth the mass of a carbon-12 atom is called *atomic mass unit* (amu) and is $= 1.66033 \times 10^{-27}$ kg.

Relative molecular mass or *molecular weight* of an element or compound is the ratio of the average mass per molecule of its naturally occurring form to

1/12th the mass of carbon-12 atom and equals the sum of the relative atomic masses of the atoms comprising it.

1.3.5 Mechanics

Mechanics is the branch of physics that deals with forces acting on bodies and the resulting motions. The subject started with Newtonian mechanics that described the translational and rotational motion of bodies under Newton's three laws of motion, but it was extended to liquids and gases (fluids) as fluid mechanics, molecular motion as statistical mechanics, and subatomic particles as quantum mechanics. A *quantum* is a discrete minimum amount by which a physical quantity changes.

Position of a body is specified by its (*x*, *y*, *z*) coordinates in three dimensions. *Displacement* of a body is the distance traveled by it in a given direction. *Velocity* of a body is the distance traveled per unit of time in a given direction. *Mass* (*m*) of a body is the amount of matter contained in it, defined in terms of its opposition to acceleration. *Acceleration* of a body is the rate of change of its velocity. *Momentum* of a body is the product of its mass and velocity. *Force* (*F*) is any agent that changes the speed or direction of motion of a body.

$$\text{Force} = \text{Mass of the body} \times \text{Acceleration}.$$

Density of a substance is its mass per unit volume. *Viscosity* of a fluid is the force acting per unit area between its layers per unit velocity gradient.

Pressure is the force applied per unit area on a body. *Stress* is the force acting per unit cross-sectional area on a body tending to cause its deformation. *Strain* is the ratio of change in dimension of a body (e.g., change in length) to its original dimension (original length). The ratio of stress applied to a material to the strain induced in it is called the *elastic modulus* (*E*) of the material.

Energy (*E*) is the capacity of a physical system to perform work. *Power* is the rate of doing work. *Potential energy* is the ability of a body to do work by virtue of its position, relative to others, for example, a heavy stone on the top of a mountain possesses gravitational potential energy due to its height above the ground. *Kinetic energy* is the energy possessed by a body due to its motion, such as the energy of a moving train.

Lagrangian of a system (symbol *L*) is a function $L = T - U$ where *T* is the kinetic energy and *U* the potential energy. *Hamiltonian* of a system (symbol *H*) is a function used to express its energy in terms of momentum and positional coordinates; in simple cases, it is the sum of kinetic and potential energies.

A *wave* is a periodic disturbance moving in space or in a medium. Waves are of two types: *longitudinal* (in which the disturbance is in the same direction as that of wave propagation) and *transverse* (in which the disturbance

is perpendicular to the direction of the wave). *Wavelength* of a wave is the distance between successive compressions and rarefactions of a longitudinal wave (or crests and troughs of a transverse wave). *Frequency* of a wave is the number of cycles per second, usually expressed in Hertz (Hz). *Period* is the reciprocal of frequency. *Amplitude* is the largest difference of the disturbed quantity from its mean value.

Two fundamental differences exist between quantum and classical mechanics:

1. The parameters of a quantum system such as the position and momentum of an electron are affected by the act of measuring them. This is expressed in the *Uncertainty principle*, which states that the uncertainty in simultaneous measurement of *x*-coordinate and *x*-component of momentum of the particle

$$\Delta x \Delta p_x \geq \frac{h}{4\pi}$$ (1.1)

   where *h* is the Planck's constant.

2. The parameters of a quantum system are not exact or precise values but are expressed as probability distributions. This happens because electrons act as particles and waves at the same time; this concept is called *wave–particle duality*. The wavelength associated with a particle of mass *m* moving with a velocity *v* is

$$\lambda = \frac{h}{mv}$$ (1.2)

For an electron in an atom, the ideas of *atomic and molecular orbitals* originate from here (see Section 1.4.9).

Wave function or *eigenfunction* $\psi(x, y, z)$ of a particle is a mathematical expression involving the coordinates of the particle in space such that the square of its absolute value $|\psi|^2$ at a point is proportional to the probability of finding the particle in a small elementary volume *dxdydz* at that point. Each permitted wave function of a particle has a corresponding allowed energy level called the *eigenvalue*. The wave functions are the solutions of the *Schrödinger's equation* whose time-independent form is

$$\nabla^2 \psi + \frac{8\pi^2 m(E - U)\psi}{h^2} = 0$$ (1.3)

where
  m is the mass of the particle
  E is its total energy
  U is the potential energy

1.3.6 Heat

Heat or *thermal energy* is the energy of vibration, rotation, or translation of the atoms or molecules of a body. It is transferred from one place to another by three modes: (i) *conduction* (through successive collisions from high-kinetic-energy molecules to neighboring low-kinetic-energy molecules), (ii) *convection* (by mass movement of high-kinetic-energy molecules from a high-temperature region to a low-temperature region), and (iii) *radiation* (by electromagnetic waves at the speed of light, not requiring any material medium).

Temperature is a measure of the degree of hotness or coldness of a body, that is, the average kinetic energy of the molecules of the body. *Absolute zero* is the theoretical lowest attainable temperature $= -273.15°C$ or $0K$, at which all atomic motion has ceased. It cannot be reached because it requires an infinite amount of energy.

Specific heat capacity of a substance is the amount of heat required to raise the temperature of unit mass of the substance by $1°C$. *Latent heat* is the quantity of heat evolved or absorbed to change the state of a unit mass of a substance without any change in its temperature.

Heat is measured in kilocalories (kcal). The amount of heat required to raise the temperature of $1\,kg$ of the substance by $1°C$ is $1\,kcal$. Heat $=$ mass $\times$ specific heat $\times$ rise in temperature.

1.3.7 Sound

Sound or *acoustic energy* is the energy of a vibrating object that is transmitted through a medium in the form of longitudinal pressure waves consisting of compressions (high-pressure regions) and rarefactions (low-pressure regions) taking place in the same direction as that of wave propagation.

Sound requires a material medium for propagation, and its speed is different in various media. The speed of sound in a medium is the square root of the ratio of elastic bulk modulus of the medium to its density. In air, the speed of sound is $330\,m\ s^{-1}$. *Audible sound* has frequency between 20 and 200,000 Hz. *Infrasound* has frequency below 20 Hz. *Ultrasound* has frequency above 20,000 Hz. Ultrasonic waves are used for nondestructive flaw detection in metals. They are also used in medical diagnosis in pregnancy where x-rays could be harmful.

The *decibel* (symbol dB) is a dimensionless unit of power ratio used in acoustics. For a pair of powers P_2 and P_1, the power ratio in decibels is $10 \log_{10} (P_2/P_1)$.

1.3.8 Light

Light energy consists of visible electromagnetic radiations with wavelengths from 400 to 750 nm. Light is propagated as waves at a speed of $3 \times 10^8\ m\ s^{-1}$ in vacuum. It is also conveyed as discrete energy packets or quanta called *photons*.

Reflection is the process in which light bounces back or is deflected at the boundary between two media. *Refraction* is the process of change of direction in which light bends in crossing from one medium to another of different density. The ratio of speed of light in vacuum to speed of light in a medium is called the *refractive index* of the medium. *Dispersion* is the splitting of light of mixed wavelengths into its component colors. *Diffraction* is the spreading of light when it passes through an aperture or around the edge of a barrier. *Polarization* is the process of confining the vibrations of the electric vector of light waves in one direction; in unpolarized light, the electric field vibrates in all directions perpendicular to the direction of propagation.

Fluorescence is the emission of light by a material on irradiation with electromagnetic waves. It lasts for a few microseconds. *Phosphorescence* is a similar phenomenon lasting for a few minutes.

1.3.9 Electricity

Electrical energy is a form of energy arising from the existence of charged bodies. A body is said to be *electrically charged* if on rubbing with another body it acquires the ability to attract light objects like pieces of paper, fur, etc. The charge produced on a glass rod rubbed with silk is called *positive charge* while that created on an ebonite rod rubbed with flannel is known as *negative charge. Unlike charges attract each other and like charges repel. Electric field* is the region of space in which force is exerted by the charge. *Electric field strength* or *intensity* (E) at a point in an electric field is defined as the force per unit charge experienced by a small charge placed at that point. *Electric current* (I) is the flow of electric charge and its magnitude is given by the rate of flow of charge, that is, the amount of charge per unit time. *Circuit* is the closed path around which electric current flows. *Electrical potential* (V) at a point in an electric field is the work done in transferring a unit positive charge from infinity to that point while *potential difference between two points* (ΔV) is the work done in transferring a unit positive charge from one point to the other.

Electric displacement is the displacement of charge produced in an electric field per unit area perpendicular to the direction of the field. *Permittivity* is the ratio of electric displacement to electric field. *Relative permittivity* or *dielectric constant* of a material is its permittivity with respect to permittivity of free space. *Displacement current* is the rate of displacement of charge in an insulator (a material that does not allow current to flow through).

Steady current of constant magnitude flowing in one direction is called *direct current* (DC). *Alternating current* (AC) is a current that periodically changes its magnitude and direction.

Resistance of a body is its opposition to the flow of electric current. *Conductance* is the reciprocal of resistance. *Resistivity* of a material is the resistance of a unit cube of the material. *Conductivity* is the reciprocal of resistivity. *Resistor* is a circuit component designed to give a specified resistance value in a circuit.

Capacitance of a conductor is the charge produced on it per unit potential applied. *Capacitor* is a circuit component for storing electric charge, designed to provide a fixed or variable value of capacitance in a circuit. A *parallel plate capacitor* consists of two metal plates separated by an insulator.

Inductance is the property by which an EMF is generated in a conductor due to a changing current in itself (self inductance) or a neighboring circuit (mutual inductance). *Inductor* is a circuit component fabricated to give a specified inductance value. It generally consists of a coil wound around a core of magnetic material.

Impedance in a circuit is the sum of contributions from resistors, capacitors (in the form of capacitive reactance $X_C = 1/\omega C$), and inductors (as inductive reactance $X_L = \omega L$).

1.3.10 Magnetism

Magnetic energy is the energy associated with magnets. A *magnet* is a piece of iron, steel, alloy, ore, etc., showing the property of attracting iron or similar materials called *magnetic materials. Magnetic field* is the region surrounding a magnetic pole in which the magnetic force due to it is perceived. A *magnetic pole* is each of the two regions of a magnet from which the magnetic force appears to originate. The strength and direction of magnetic field (H) is expressed in terms of the magnetic flux density or magnetic induction, symbol B, defined as the magnetic flux per unit area of a magnetic field perpendicular to the magnetic force. *Flux* is a measure of the quantity of magnetism taking into account the strength and extent of the magnetic field. *Magnetic permeability* is the ratio of magnetic flux density to magnetizing field.

Magnetism is the study of magnetic phenomena and their laws. Magnetism arises from the spinning motion of electrons so that each electron produces a small magnetic field. The magnetic effects of electrons spinning in opposite directions cancel out. Magnetism is of four types: diamagnetism, paramagnetism, ferromagnetism, and ferrimagnetism. In *diamagnetism*, the magnetization is opposite to the applied magnetic field, weak and temporary. Diamagnetic materials contain paired electrons. In *paramagnetism*, it is in the same direction as the applied magnetic field, but weak and temporary. In *ferromagnetism* occurring in materials like iron, cobalt, and nickel, there is an enormous increase in magnetization in the same direction as the field due to the alignment of regions of aligned electron spin called domains. Moreover, magnetism is retained even after removal of the field. Both paramagnetic and ferromagnetic materials contain unpaired electrons. *Antiferromagnetism* is a property possessed by some metals, alloys, and salts of transition elements such as manganese oxide (MnO) by which the atomic magnetic moments form an ordered array that alternates or spirals so as to give no net total moment in zero applied magnetic field and hence almost no gross external

magnetism. In *ferrimagnetism*, which is observed in composite materials such as magnetite, not individual elements, the overall spin effect in one direction is greater than that in the other.

1.3.11 Electromagnetism

Scottish physicist J.C. Maxwell formulated four basic equations, popularly known as *Maxwell equations*, unifying electricity and magnetism by describing the space and time dependence of electric and magnetic fields: (i) a relationship between electric field and electric charge, (ii) a relationship between magnetic field and magnetic poles, (iii) generation of an electric current by a changing magnetic field, and (iv) production of a magnetic field by an electric current or a changing electric field.

   Electromagnetic radiation is the energy emitted by an accelerated charge and its accompanying electric and magnetic fields. An *electromagnetic wave* consists of oscillating electric and magnetic fields at right angles to each other and the direction of propagation. *Electromagnetic spectrum* is the range of wavelengths or frequencies over which the electromagnetic waves extend. It covers radio waves, infrared (IR) waves, visible light, ultraviolet waves, x-rays, and gamma rays in the descending order of wavelengths.

1.3.12 SI System of Units

SI (Systéme International d' Unités) is a coherent system of metric units for scientific usage across the world, based on seven basic units: meter (length), kilogram (mass), second (time), ampere (electric current), Kelvin (thermodynamic temperature), mole (amount of substance), and candela (luminous intensity). *Meter* (m) is the length of path traveled by light in a vacuum during an interval of $1/(2.99\ 792\ 458 \times 10^8 \text{ s})$. *Kilogram* (kg) is the mass of the platinum–iridium (PtIr) prototype kept by the International Bureau of Weights and Measures at Sévres near Paris, France. *Second* (s) is the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of ground state of Cesium-133 atom. *Ampere* (A) is the constant current that when maintained in two straight, parallel, infinitely long conductors of negligible cross section will produce a force of $2 \times 10^7 \text{ N m}^{-2}$ between the conductors. *Kelvin* (K) is the temperature $= 1/273.16$ fraction of the thermodynamic temperature of triple point of water, the temperature and pressure at which vapor, liquid, and solid phases of water are in equilibrium $= 0.01°C$ or $273.16\,K$ and $611.2\,Pa$. *Mole* (mol) is the amount of substance that contains as many elementary units (atoms, molecules, ions, radicals, electrons, etc.) as there are atoms in $0.012\,kg$ of Carbon-12, an isotope of carbon. *Candela* (Cd) is the luminous intensity in a given direction of a source that emits monochromatic radiation at a frequency of $540 \times 10^{12} \text{ Hz}$ and has a radiant intensity in this direction of $1/683\,W\ sr^{-1}$.

1.4 Chemistry

1.4.1 Definition of Chemistry

Chemistry is the science of the properties, structure, and composition of substances, and how these undergo transformations (Tro 2010).

1.4.2 Elements and Compounds

Pure substances have an invariable composition and are composed of either elements or compounds. *Elements* are substances that cannot be decomposed into simpler substances by chemical means. *Compounds* can be decomposed into two or more elements. They are substances made by the combination of two or more elements in fixed proportions.

1.4.3 Organic and Inorganic Compounds

Organic compounds are those that contain carbon. They contain one or more of the following elements: carbon (C), hydrogen (H), oxygen (O), nitrogen (N), sulfur (S), halogens (fluorine [F], chlorine [Cl], bromine [Br], iodine [I], and astatine [At]), and occasionally phosphorous (P) and metals. Because of the ability of carbon atoms to form long chains, the number of such compounds far exceeds that of other elements. These compounds form the basis of living matter.

   Inorganic compounds are those that do not contain carbon and usually occur outside living organisms. These compounds are formed from elements other than carbon.

1.4.4 Subdivisions of Chemistry

What are the main subdivisions of chemistry? Physical, organic, and inorganic. *Physical chemistry* deals with the physical principles governing the chemical phenomena, the physical structure of chemical compounds, and the physical effects of chemical reactions. *Organic and inorganic chemistry* are, respectively, the branches of chemistry concerned with organic and inorganic compounds. Organic chemistry is the chemistry of the element carbon, traditionally excluding its oxides and the carbonates, which are studied in inorganic chemistry.

1.4.5 Natural and Artificial Elements

Ninety-four elements from atomic number 1–94 occur in nature. Elements heavier than plutonium (symbol: Pu, atomic number 94), the heaviest naturally occurring element, are called *artificial elements*. Elements up to atomic

number 118 have been made in laboratory. According to theoretical predictions, the highest atomic number lies between 170 and 210.

Hydrogen is the most abundant element in the universe (90.7% by mass) while oxygen is the most abundant element in the earth's crust (47.4% by mass). Tungsten (W) has the highest melting and boiling points (M.P. = 3407°C and B.P. = 5657°C).

1.4.6 Metals, Nonmetals, and Metalloids

Elements are divided into three classes: metals, nonmetals, and metalloids. *Metals* form positively charged ions (cations) by the loss of electrons from their atoms; they have positive valency. *Nonmetals* form negatively charged ions (anions) by gain of electrons by their atoms; they have negative valency. Metals (except mercury, Hg, which is a liquid) are solid at room temperature. In general, metals, for example, gold (Au), silver (Ag), copper (Cu), aluminum (Al), etc., are lustrous (shiny), malleable (can be shaped by hammering), ductile (capable of being drawn into a wire, not brittle, moldable), and good conductors of heat and electricity. Examples of nonmetals are argon (Ar), bromine (Br), carbon (C), chlorine (Cl), fluorine (F), helium (He), iodine (I), krypton (Kr), neon (Ne), nitrogen (N), oxygen (O), phosphorous (P), radon (Rn), selenium (Se), sulfur (S), and xenon (Xe). Nonmetals are generally solids or gases except bromine, which is a liquid.

Elements such as arsenic, antimony, hydrogen, silicon, germanium, and tellurium exhibit the characteristics of both metals and nonmetals, that is, properties intermediate between those of metals and nonmetals, and are called *metalloids*.

1.4.7 Periodic Table of Elements

Periodic table is a compilation of the chemical elements in ascending order of their atomic numbers to show the similarities of properties of elements. The vertical columns of the periodic table numbered I–VIII are called *groups* while the horizontal rows 1–7 are known as *periods*. The groups contain elements that are similar to each other.

1.4.8 Chemical Change and Reaction

When water is boiled, it is converted into steam. This is a physical change because only the state has changed from liquid to gas, and can be reversed by cooling when water vapor condenses. Steam and water have the same composition consisting of H_2O molecules. But when a piece of sodium (Na) is added to water, hydrogen (H_2) gas is evolved and sodium hydroxide (NaOH) is formed. This is a chemical change in which the nature of substances is altered.

A *chemical reaction* is the interaction between one or more elements or compounds called *reactants* to form new elements or compounds termed *products*. A *chemical equation* is the symbolic representation of a chemical reaction, for example, potassium chlorate decomposes on heating in presence of manganese dioxide powder into potassium chloride and oxygen. Manganese dioxide acts as a catalyst, a substance that increases the speed of a reaction without any changes in itself. Symbolically,

$$2KClO_3 = 2KCl + 3O_2 \ (MnO_2 \ catalyst) \tag{1.4}$$
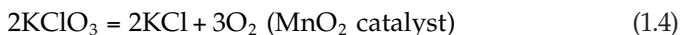
Enthalpy (*H*) is the heat content of a substance. *Entropy* (*S*) of a substance is a property related to the degree of disorder or randomness in it. *Free energy* (*F*) is the available energy of a chemical reaction. *Gibb's free energy G* is given by

$$G = H - TS \tag{1.5}$$

where *T* is the absolute temperature.

1.4.9 Electronic Configuration (Structure) of Elements

It is the arrangement of electrons in the different orbitals or shells of an atom of the element. Orbital or shell is the region in space surrounding the nucleus in which there is highest probability of finding the electron. For example, the electronic structure of oxygen is $1s^2 \ 2s^2 \ 2p^4$. The symbols s, p, d, f, … represent the sub-shells. The prefixes 1, 2, … denote the shell number while the superscript is the number of electrons in a sub-shell.

1.4.10 Chemical Bond

It is the attractive force between the atoms in a molecule that binds together the atoms. It originates from the tendency of atoms to acquire the stable inert gas configuration. The main types of bonds are as follows: (i) *ionic or electrovalent bond* in which electrons are transferred from one atom to another forming positive and negative ions that stick together by electrostatic force; (ii) *covalent bond* in which the atoms are held together by sharing electron pairs; bonds formed by sharing one, two, and three pairs of electrons are called *single, double, and triple bonds* respectively; (iii) *hydrogen bond* in which hydrogen atom attached to one of the three elements (fluorine, oxygen, or nitrogen) is able to form a bridge with another one of these three elements; and (iv) *metallic bond*, which forms in metals, which, due to their low ionization energies, lose one or more of their outer shell electrons becoming positive ions so that the metal is pictured as a sea of free mobile electrons in which positive ions are immersed. The free electrons are said to be delocalized. *Delocalization* is the spreading of electrons in a molecule over the molecule.

1.4.11 Oxidation and Reduction

Oxidation of an atom, molecule, or ion is a reaction in which there is a loss of electrons from it, for example, in the conversion of magnesium (Mg) into magnesium oxide (MgO), Mg^{2+} and O^{2-} ions are formed. As Mg^{2+} ion is obtained from Mg by the loss of two electrons ($2e^{-}$), magnesium is oxidized here. *Reduction* of an atom, molecule, or ion is the reaction in which a gain of electrons takes place by it. The overall reaction is called reduction–oxidation or *redox reaction*.

1.4.12 Acid, Base, and Salt

An *acid* is a substance that dissociates in aqueous solution to give hydrogen ions, for example, HCl and H_2SO_4. It acts as a proton donor. It is a substance that has a tendency to accept an electron pair to form a covalent bond.

A *base* is a substance that dissociates in aqueous solution to give hydroxide ions, for example, KOH and NaOH. It acts a proton acceptor. It is a substance that can donate an electron pair to form a covalent bond. An *alkali* is a base that is soluble in water.

A *salt* is the product of the neutralization reaction between an acid and a base formed by replacing the hydrogen of an acid with a metal or electropositive ion such as the ammonium ion. An example is the sodium chloride (NaCl) salt formed by the reaction between HCl and NaOH. Similarly, Na_2SO_4 is produced from H_2SO_4.

1.4.13 Expressing Concentrations of Solutions and Gases

Concentration of a solution is the number of grams of solute per liter of solution or number of moles per liter (molar concentration). Concentration of a gas in a mixture of gases is the percentage (parts per hundred) of the gas in the mixture. It is also expressed in parts per million ($1/10^6$), ppm; or parts per billion ($1/10^9$), ppb.

1.4.14 Hydrocarbons: Saturated and Unsaturated

These are compounds made of two elements carbon and hydrogen only. Hydrocarbons are classified as follows: (i) *aromatic hydrocarbons* such as benzene (Figure 1.2) in which carbon atoms are arranged in an unsaturated ring with delocalized electrons above and below and (ii) *aliphatic hydrocarbons* in which carbon atoms are arranged in an open chain with no delocalized electrons; these are further subdivided into alkanes, alkenes, alkynes, and alicyclic compounds.

Alkanes (e.g., methane (CH_4), ethane (C_2H_6), and propane (C_3H_8)) are hydrocarbons with the general formula C_nH_{2n+2}. *Alkenes* (e.g., ethane (C_2H_4), propene (C_3H_6), and butene (C_4H_8)) are hydrocarbons having the general
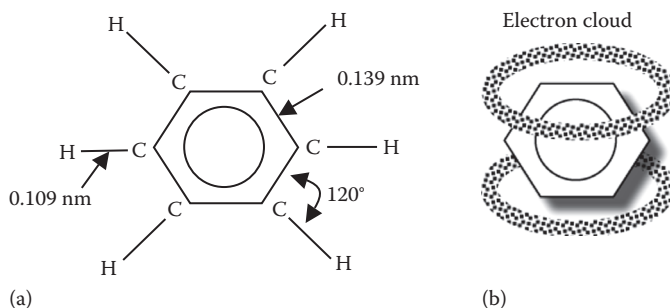
FIGURE 1.2
Benzene ring: (a) conventional symbol and (b) delocalization of electrons causing high electron
density above and below the ring.

formula C_nH_{2n}. *Alkynes* (e.g., ethyne (C_2H_2), propyne (C_3H_4), and butyne
(C_4H_6)) are hydrocarbons with the general formula C_nH_{2n-2}. *Alicyclic com-
pounds* (e.g., cyclobutane [$(CH_2)_4$], cyclopentane (C_5H_{10}), and cyclohexane
(C_6H_{12})) are the class of cyclic hydrocarbons that closely resemble open
chain hydrocarbons.

Organic compounds containing single bonds are said to be *saturated com-
pounds*; those containing multiple bonds are *unsaturated*. As alkanes have sin-
gle carbon-to-carbon atom bonds, they are saturated hydrocarbons. Alkenes
having double carbon–carbon bonds and alkynes with triple carbon–carbon
bonds are unsaturated hydrocarbons.

1.4.15 Alkyl and Aryl Groups

Alkyl group (symbol R) is a group formed from an alkane by removal of a
single atom of hydrogen ($-CH_3$, methyl group; $-CH_3CH_2$, ethyl group, etc.).
Aryl group is a group formed from an aromatic compound by removing a
single hydrogen atom, such as the C_6H_5- group from benzene.

1.4.16 Alcohols and Phenols

Alcohol is one of a class of organic molecules containing an alkyl group con-
nected to a hydroxyl group, ROH, for example, methyl alcohol (CH_3OH) and
ethyl alcohol (CH_3CH_2OH). Phenol is one of a family of organic compounds
containing an aryl group attached to the hydroxyl group, for example,
C_6H_5OH, hydroxybenzene.

1.4.17 Carboxylic Acids

These are weak organic acids containing the group $-COOH$, for example,
$HCOOH$, methanoic or formic acid; CH_3COOH, ethanoic or acetic acid.

1.4.18 Aldehydes and Ketones

Aldehydes are organic compounds containing the alkyl group, aryl group, or hydrogen attached to a carbonyl group, $=CO$, with an adjacent hydrogen, that is, $-CHO$, for example, methanal (formaldehyde), $HCHO$; ethanal (acetaldehyde), CH_3CHO.

Ketones are organic molecules containing a carbonyl group, $=CO$, attached to two alkyl groups, with the structure: $R-CO-R$, for example, CH_3COCH_3, propanone.

1.4.19 Amines and Amino Acids

These are alkyl or aryl derivatives of ammonia (NH_3), with the structure: $R-NH_2$, for example, CH_3NH_2, methylamine; $C_2H_5NH_2$, ethylamine; and $C_6H_5NH_2$, phenylamine. The group $-NH_2$ is called the amino group.

Amino acids are compounds containing both an acidic carboxyl group $-COOH$ and a basic amino group $-NH_2$.

1.4.20 Lipids

These are water-insoluble biomolecules such as fats, oils, and waxes. Fats are organic compounds that are made up of carbon, hydrogen, and oxygen. They are a source of energy in foods. Saturated fats are found in animal products such as butter, cheese, whole milk, ice cream, and fatty meats. Unsaturated fats are almost always plant-based. Mostly, but not all, liquid vegetable oils are unsaturated fats. Fats usually imply substances that are solids at normal room temperature while oils generally refer to fats that are liquids at normal room temperature. Oils also mean substances that do not mix with water and have a greasy feel, such as essential oils. Hydrogenated oils are oils that are hardened such as hard butter and margarine. Waxes are various natural, oily, or greasy heat-sensitive substances such as beeswax.

1.4.21 Carbohydrates

These are a class of biological molecules composed of carbon with hydrogen and oxygen, in the ratio of two O atoms and one H atom per C atom. Carbohydrates are of three types: monosaccharides ($C_x(H_2O)_y$, where x is at least 3); disaccharides; and polysaccharides. Monosaccharides are simple sugars like glucose (blood sugar) and fructose (fruit sugar). Disaccharides and polysaccharides contain two and multiple units of monosaccharides, respectively. Examples of disaccharides are sucrose (cane sugar), lactose (milk sugar), and maltose (malt sugar), while those of polysaccharides are starch and cellulose.

1.4.22 Proteins and Enzymes

Proteins are polymers made by linking together small molecules called amino acids. They contain carbon, hydrogen, oxygen, nitrogen, and sometimes

phosphorous and sulfur. Enzymes are proteins produced by cells that act as catalysts in living beings serving to accelerate biochemical reactions without themselves being altered.

1.5 Biology

1.5.1 What Is Biology?

Biology is the science of life and of living organisms, including their processes and phenomena (Campbell and Reece 2004). *Organism* is an individual form of life, capable of growing and reproducing.

1.5.2 Subbranches of Biology

What are the main subbranches of biology? Botany and zoology. *Botany* is the scientific study of all aspects of plant kingdom. *Zoology* deals with all aspects of the animal kingdom.

1.5.3 Origin and Evolution of Life

According to modern theory, life began from simple chemical compounds such as water (H_2O), methane (CH_4), and ammonia (NH_3), which were present in the original atmosphere. Lightning fused these chemicals into amino acids, the building blocks of life. The production of organic compounds from such simple inorganic compounds was possible in the early hydrogen-rich atmosphere. The organic compounds fell from the atmosphere onto the earth's surface where the rain water washed them into pools and ultimately into oceans. Gradually, the organic molecules joined together into long chains, proteins, and deoxyribonucleic acids (DNAs) until a cell appeared that could replicate itself.

   Evolution is the process by which living organisms have progressed from simple ancestral forms to present complex forms.

1.5.4 The Cell

The *cell* is the basic structural and functional unit of all organisms (Figure 1.3). The *protoplasm* is the living part of a cell consisting of a nucleus embedded in a membrane-enclosed *cytoplasm*, which is the protoplasmic content excluding the nucleus. The *nucleus*, which houses DNA (DNA: the double-stranded, helical molecular chain carrying genetic information), is contained within a membrane and separated from other cellular structures. Its functions are cell regulation and reproduction. *Chromatin* consists of masses of DNA and
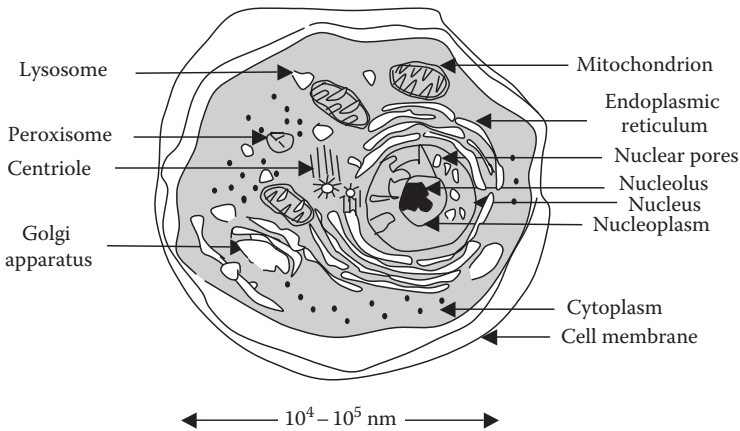
FIGURE 1.3

Generalized structure of animal cell. The cell is surrounded by a membrane made up of two fluid layers of fatty molecules. The cytoplasm comprises the total cell contents including the subcellular structures. The nucleus contains the genetic information. Mitochondria provide energy in a useful form through respiration. The endoplasmic reticulum is the site of protein and lipid production. The Golgi apparatus modifies proteins and lipids and distributes them to the rest of the cell.

associated proteins. *Chromosome* is a rod-like structure in the nucleus of the cell that carries the *genes* (segments of DNA that are the basic units of heredity) of cells and performs an important role in cell division and transmission of hereditary characters. Nucleus contains a small body *nucleolus* in which RNA (ribonucleic acid) is made. *Mitochondria* provide energy to the cell by carrying out respiration. *Lysosomes* are dark spherical bodies in the cytoplasm, which contain enzymes that break down complex compounds into simpler subunits. *Endoplasmic reticulum* (ER) is the site of protein and lipid production. *Golgi apparatus* modifies proteins and lipids, and distributes them to the rest of the cell. *Plastids* are hollow spherical bodies (leucoplasts contain starch; chloroplasts contain chlorophyll). *Vacuoles* are fluid-filled sacs. *Microtubules* are tiny tubes associated with transport inside cells. *Centrioles* are cylindrical bodies concerned with cell division and movement. *Cilia and flagella* are motile hairs projecting from cells, and linked to movements of cells and substances.

What are the basic types of cells? There are two basic types of cells: those with nuclei (*eukaryotic*) and those without nuclei (*prokaryotic*). All of the cells in a human being are eukaryotic.

1.5.5 Differences between Bacteria and Viruses

Two common terms are "bacteria" and "viruses." Bacteria (singular: *bacterium*) are large unicellular microorganisms, typically a few micrometers long

and having many shapes including curved rods, spheres, rods, and spirals. Viruses (from the Latin noun *virus*, meaning toxin or poison) are submicroscopic particles (ranging in size from 20 to 300 nm) that can infect the cells of a biological organism.

Bacteria and viruses are clearly distinguishable as follows: (i) Viruses are the smallest and simplest life form known. They are 10–100 times smaller than bacteria. (ii) Viruses must have a living host, like a plant or animal, to multiply, while most bacteria can grow on nonliving surfaces. (iii) Bacteria are intercellular organisms, that is, they live in-between cells; whereas viruses are intracellular organisms, that is, they infiltrate the host cell and live inside the cell. (iv) Some bacteria are useful but all viruses are harmful. (v) Antibiotics can kill bacteria but not viruses. (vi) An example of a disease caused by bacteria is *strep throat*. An affliction caused by a virus is the *flu*.

1.5.6 Heredity, Chromosomes, Genes, and Related Terms

Heredity is the transmission of characteristics from one generation to successive generations. *Chromosome* is the thread-like in the nucleus of eukaryotic cells composed of DNA and protein.

Gene, the fundamental unit of inheritance located in the chromosome, is a specific section of DNA that codes for a recognizable cellular product, either RNA or polypeptide (a long >20 chain of amino acids). *Genetic code* is the information encoded as a linear sequence of four nitrogenous bases that form the structure of DNA. *Genome* is the full set of chromosomes of an individual or the total number of genes in such a set. *Genetics* is the study of heredity and the gene.

1.6 Semiconductor Electronics

1.6.1 What Is Semiconductor Electronics?

Electronics deals with the control of movement of electrons in solid-state materials, vacuum, and gases to perform specialized tasks for computing, power conversion, regulation, communications, etc. *Semiconductors* are materials having conductivities (10^5–10^{-7} S m^{-1}) intermediate between metals (as high as 10^9 S m^{-1}) and insulators (as low as 10^{-15} S m^{-1}). *Semiconductor electronics* is the portion of electronics dealing with the use of semiconductors for controlling the flow of electrons.

1.6.2 Energy Bands in Conductors, Semiconductors, and Insulators

When isolated atoms come together to form solids, the discrete energy levels of atoms split up into multiple levels. Due to this splitting, solids are
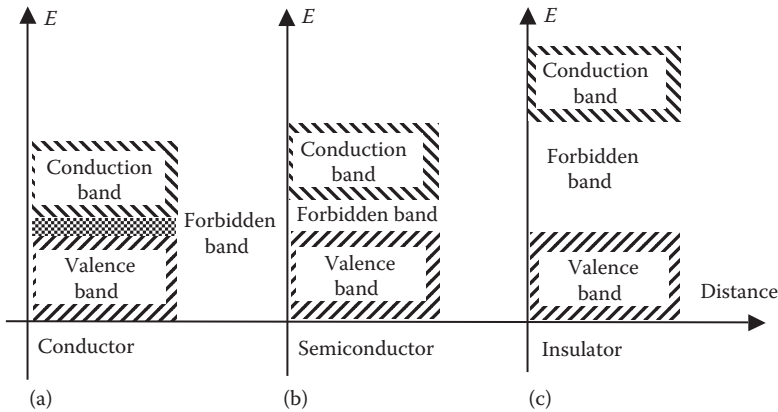
FIGURE 1.4
Energy band diagrams for: (a) conductor, (b) semiconductor, and (c) insulator.

described by energy bands instead of energy levels. Electrons in valency shells of solids are said to reside in the valence band while the electrons that acquire energy by some means such as heat, light, etc., are dislodged from their chemical bonds and become free to take part in electrical conduction. These are in conduction band.

Energy band diagram is a plot of energy of electrons with respect to distance through the solid (Figure 1.4). It contains two bands: valence and conduction bands. The energy difference between the top of the valence band and bottom of the conduction band is called the *forbidden energy gap* or *band gap*.

In metals, the energy bands either overlap each other or a very small band gap exists between conduction and valence bands. This means that approximately all the electrons lie in the conduction band and contribute to conduction. Semiconductors have a medium energy gap while insulators have a large bandgap. Popular semiconductors: germanium (Ge), silicon (Si), and gallium arsenide (GaAs) have bandgap values of 0.66, 1.1, and 1.42 eV (electron volt), respectively.

One *electron volt* is the energy gained by an electron when it is accelerated by a potential difference of 1 V (=potential difference between two points of a conductor carrying a current of 1 A when the power dissipation is 1 W). Insulators like silicon dioxide (SiO_2) and silicon nitride (Si_3N_4) have bandgaps of 9 and 5 eV, respectively.

1.6.3 Interesting Properties of Semiconductors

What are the special properties of semiconductors that make their study important for us? The electrical conductivity of semiconductors can be altered by the controlled addition of small amount of impurities. This controlled addition of impurities to a pure semiconductor material called intrinsic semiconductor is known as *doping*.
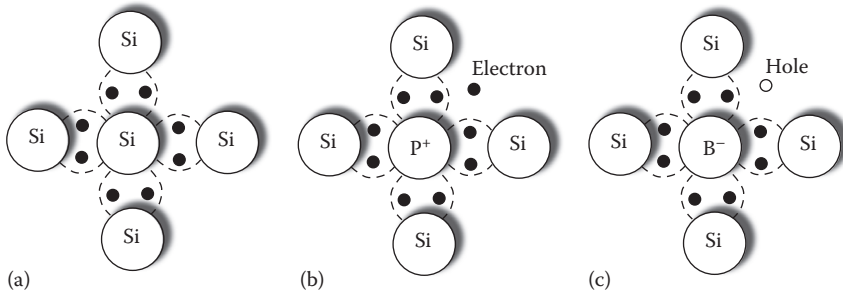
FIGURE 1.5
Bond diagrams: (a) intrinsic semiconductor, (b) N-type semiconductor, and (c) P-type semiconductor.

Silicon used for device fabrication is in a *single crystal form*. A crystal consists of an ordered arrangement of atoms arranged at fixed distances repeated throughout the material in a periodic fashion.

Silicon (Si: valency 4) has four electrons in its outermost shell. Every atom in a single crystal is surrounded by four atoms bound to it by covalent bonds formed by sharing of electrons (Figure 1.5a). Each shared electron pair consists of one electron from the central silicon atom and one electron from the surrounding silicon atom. Thus, four pairs of electrons add up to eight electrons completing the stable inert gas configuration.

Suppose an intrinsic silicon crystal is doped with a *pentavalent impurity* such as phosphorous (P), arsenic (As), or antimony (Sb) (Figure 1.5b). When a silicon atom in the crystal is substituted by a phosphorous atom, the four electrons of the phosphorous atom take part in covalent bonding as in silicon but the fifth electron is available for conduction. Thus, one free electron is donated by each phosphorous atom. Phosphorous is therefore called a *donor impurity* and because of the resulting negatively charged electrons present in a phosphorous-doped silicon crystal, the material is referred to as an *N-type semiconductor*.

If doping is done by a trivalent impurity like boron (B) (Figure 1.5c), indium (In), or gallium (Ga), three bonds around the impurity atom are complete but the fourth bond has one missing electron. An electron from an adjacent silicon atom jumps to fill and complete this bond. The electron vacancy left behind in the silicon atom is called a *hole* and acts like a positively charged particle. As one hole is produced per impurity atom, the silicon crystal has the same number of holes as the number of impurity atoms. As these impurities accept electrons from the silicon, they are termed *acceptor impurities*. Due to the presence of a large number of positively charged holes in boron-doped silicon, it is called a *P-type semiconductor*.

In an N-type semiconductor, the number of electrons is larger than the number of holes. Electrons carry the major part of current and are called

majority carriers. Holes are termed the *minority carriers*. In a P-type semiconductor, holes are majority carriers and electrons are minority carriers.

Thus, by deliberate incorporation of known amounts of phosphorous or boron atoms in semiconductors, conductivities of resulting N-type and P-type semiconductors can be varied within wide limits. However, both N-type and P-type semiconductors are electrically neutral as a whole because the number of electrons = number of positively charged donor ions in an N-type semiconductor and the number of holes = number of negatively charged acceptor ions in a P-type semiconductor.

The distribution of electrons at different energies obeys the *Fermi–Dirac distribution function*, which expresses the probability that a state at energy E is occupied by an electron:

$$f(E) = \left[1 + \exp\left(\frac{E - E_F}{kT} \right) \right]^{-1} \tag{1.6}$$

where
E_F is the Fermi level
T is the absolute temperature

Fermi level is the energy at which there is 50% probability of being occupied by an electron. In an intrinsic semiconductor, the *Fermi level* is located approximately midway between the conduction and valence bands. When a semiconductor is doped with donor impurity, this probability increases and therefore the Fermi level shifts upward. On doping with acceptor impurity, the situation reverses and Fermi level shifts downward.

What happens when a semiconductor material is heated or when we shine light on the semiconductor? Conductivities of semiconductors are altered by the action of heat and light. Heat or light energies break the chemical bonds leading to the formation of electron–hole pairs. These electron–hole pairs participate in electrical conduction enhancing the conductivity of the material.

1.6.4 P–N Junction

At a junction between P-type and N-type semiconductors, holes diffuse from P-side to N-side and electrons diffuse from N-side to P-side (Figure 1.6). *Diffusion* is the movement of particles under a concentration gradient, that is, from a high-concentration to a low-concentration region. The above diffusion of carriers constitutes the *diffusion current* and leaves negatively charged acceptor ions on the P-side and positively charged donor ions on the N-side. Thus, an electric field is created around the junction. This electric field is directed from N- to P-side. It sets up an electric current in the reverse direction to the diffusion current flowing under the concentration gradient. This
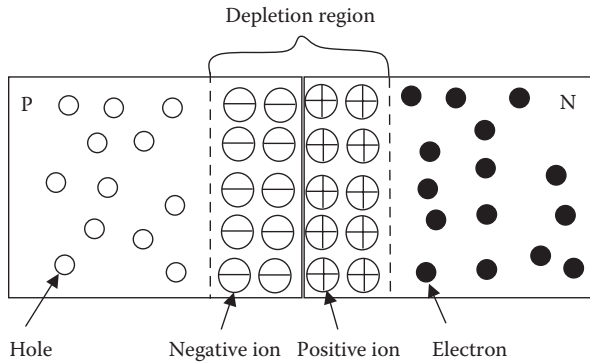
FIGURE 1.6
P–N junction diode.

current set up by the electric field is called the *drift current*. Ultimately, the number of carriers crossing the junction from one side to the other by diffusion and those crossing in reverse direction due to the electric field becomes equal. Then a *dynamic equilibrium* is established. The potential difference associated with the electric field at the junction is called the *built-in potential*.

When a power supply is connected across a P–N junction with its positive terminal on P-side and negative terminal on N-side, the P–N junction is *forward biased*. As long as the applied voltage is less than the built-in potential, there is no current flow across the junction. But no sooner than this voltage exceeds the built-in potential, the aforesaid dynamic equilibrium is disturbed and current starts flowing by diffusion. Thus, current flowing across a forward-biased diode is a *diffusion current*.

On changing the connections of the power supply with positive voltage on N-side and negative voltage on P-side, the diode is *reverse biased*. Here, the depletion region is widened and thickness of the depletion region depends on the applied voltage. Any minority carriers in the depletion region or close to it, such as electrons on P-side or holes in N-side, are carried and swept away by the applied field leading to a *reverse leakage current*. Thus, reverse current is essentially a *drift current*.

At high voltages, one of the following two phenomena takes place: (i) avalanche multiplication and (ii) punchthrough breakdown. In *avalanche multiplication*, the carriers, electrons and holes, are accelerated to high velocities. They collide with electrons in other atoms dislodging them from their orbits. In this way, an avalanche of carriers builds up by *impact ionization* producing high currents causing breakdown. In *punchthrough mechanism*, the depletion region enlarges to the full thickness of lightly doped side touching the contact. Then, a very high current starts flowing. Another breakdown mode is *Zener breakdown*, which occurs if both N- and P-sides of the junction are heavily doped. The depletion region is very thin so that carriers *tunnel* across this region causing breakdown. It usually takes place around 6 V.

1.6.5 Bipolar Junction Transistor

It is of two types: NPN and PNP. The *NPN transistor* (Figure 1.7a and b) contains three regions: a heavily doped N region (emitter), a lightly doped thin P region (base), and a still lighter doped N region (collector). On forward biasing the emitter–base junction and reverse biasing the collector–base junction, electrons are injected from the emitter into base. Very few electrons are lost by recombination in the thin base region. Practically, more than 99% of the injected electrons reach the collector junction. Although the collector current is slightly smaller than the emitter current, the important observation is that the transistor is able to pass current from a low-resistance forward-biased emitter–base junction to a high-resistance reverse-biased collector–base junction; hence the name "transfer resistor" or "transistor." The significant advantage here is that a high resistance can be connected in the collector circuit across which a large voltage is developed yielding a high voltage gain. This is *voltage amplification*. The term "amplification" means strengthening a signal by drawing power from a source other than that of the signal. Clearly, the transistor does not generate extra energy nor is the
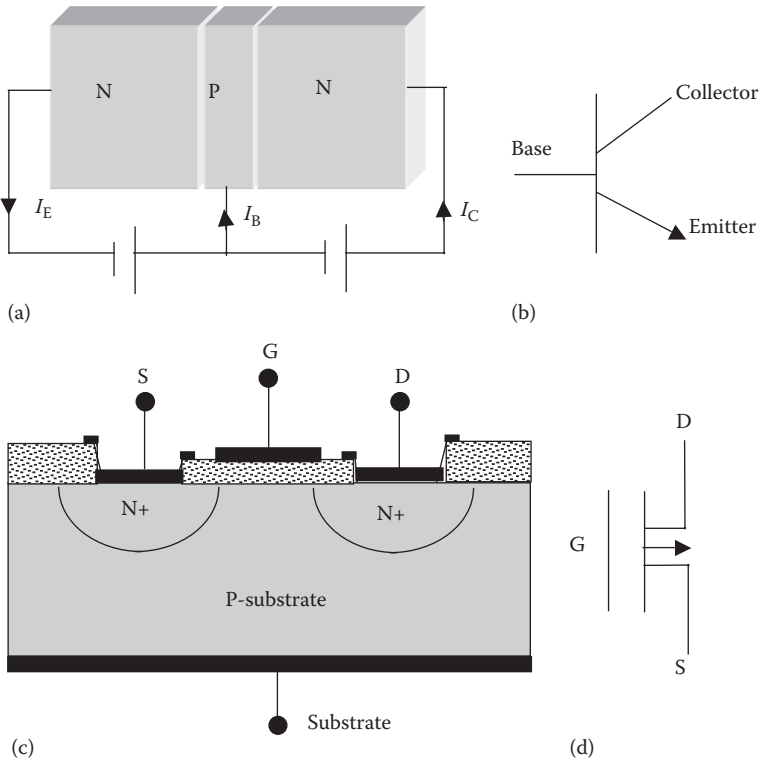
FIGURE 1.7
Transistors: (a) BJT and (b) its circuit symbol; (c) MOSFET and (d) its circuit symbol.

principle of conservation of energy violated. The extra power comes from the supplies in the circuit.

1.6.6 Metal-Oxide-Semiconductor Field-Effect Transistor

It is of two kinds: enhancement or normally off device without a built-in channel and depletion or normally on mode device having a built-in channel. The N-channel enhancement mode device (Figure 1.7c and d) contains two heavily doped N regions called source and drain in a P-substrate. The device has four terminals: source, drain, gate, and substrate. A positive voltage is applied between source and drain with source and substrate grounded. The source–substrate junction is shorted but the drain–substrate junction is reverse biased. Therefore, there is no current flow between source and drain. But when a positive voltage is applied to gate, by field effect through the underlying oxide layer, a depletion region is formed at the surface of the substrate below the gate. This is due to repulsion of holes by the positive gate voltage. On increasing the gate voltage, to a certain gate voltage called the *threshold voltage*, minority carrier electrons are attracted toward the surface of the P-substrate below the gate, forming an *inversion region* called the *channel*. This conducting channel connects the source and the drain, and current starts flowing between source and drain.

Unlike bipolar junction transistor (BJT), which is a current-controlled device, metal-oxide-semiconductor field-effect transistor (MOSFET) is a voltage-controlled device. Also, MOSFET has higher input impedance than BJT. Moreover, MOSFET is a *unipolar device* in which one type of carriers, either electrons or holes, carry current, whereas in BJT, both the electrons and holes are responsible for current flow; hence the name "bipolar."

1.6.7 Analog and Digital Circuits

An analog signal is one that varies continuously with time. A digital signal changes only at discrete levels with respect to time. Digital signals are less prone to noise. Commonly used analog circuit is the operational amplifier, which is a high gain DC amplifier with high input impedance and low output impedance. Inverting and non-inverting amplifiers, summing, differentiating, and integrating circuits are constructed using operational amplifiers.

Among digital circuits, mention may be made of logic circuits such as AND, OR, and NOT circuits. An AND circuit gives an output signal only when all its inputs A, B, C … are present. An OR circuit gives an input when any one of its input signals A or B or C or … is applied. A NOT circuit produces an output when there is no signal at the input. Combinations of NOT and AND circuits and NOT and OR circuits are called NAND and NOR circuits, respectively. By suitable combinations of logic circuits, mathematical operations like addition, multiplication, division, and subtraction are performed.

The different analog and digital circuits are fabricated on a single chip of silicon and are available as integrated circuits (ICs).

1.7 Nanometer and Appreciation of Its Magnitude

Is "Nano" an English word? What does it mean? The word "nano" is derived from Greek, meaning dwarf, a person shorter in height than normal, and standing for *one*-billionth. It is a unit of length = one billionth of a meter = 1×10^{-9} m.

How to imagine the sizes at such a small scale and think about magnitudes of such small dimensions? To visualize the nanometer scale, dimensions of familiar objects must be compared in nanometer (Table 1.1). The average height of a man is 1.7×10^9 nm. Thickness of a book page is $\sim 10^5$ nm. Diameter of the human hair is 7.5×10^4 nm. The minimum length that the human eye can see is 1×10^4 nm. So, the human eye can see the hair. Dimensions of protein molecules are in the range 1–20 nm, width of the DNA molecule is 2 nm while that of a carbon nanotube (CNT) is 1.3 nm. Diameter of a gold atom is 0.33 nm, that of a water molecule is 0.29 nm, and a hydrogen atom is 0.1 nm.

TABLE 1.1

Small and Big Things

S.No.	Object	Diameter	Reference
1.	Electron	$<10^{-18}$ m	http://hypertextbook.com/facts/2000/DannyDonohue.shtml
2.	Proton	1×10^{-15} m	Christensen (1990)
3.	Nucleus	1.6 fm (10^{-15} m) (for a proton in light hydrogen) to about 15 fm (for the heaviest atoms, such as uranium)	http://wiki.answers.com/Q/What_is_the_diameter_of_nucleus
4.	Atom	0.1–0.5 nm	hypertextbook.com/facts/MichaelPhillip.shtml
5.	Water molecule	0.282 nm	http://www.falstad.com/scale/
6.	Polio virus	30 nm	http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/V/Viruses.html
7.	Human red blood cell	7.5×10^3 nm	http://wiki.answers.com/Q/What_is_the_Diameter_of_red_blood_cell
8.	*Amoeba* cell	3×10^5 nm	http://www.answers.com/topic/amoeba
9.	Planet Earth (Equatorial)	12,756.32 km	geography.about.com/library/faq/blqzdiameter.htm
10.	Sun	1,392,000 km	http://www.daviddarling.info/encyclopedia/S/Sun.html

A CNT, also called a *buckytube,* is a large molecule of pure carbon that is long and thin and shaped like a tube, hundreds to thousands of nanometers long. It is 10^2 times stronger than steel and but has one-sixth its weight.

Example 1.1

Convert the average human height (5 ft 7 in.) into nanometers.

5 ft 7 in. = $5 \times 12 + 7$ in. = 67 in. = 67×2.54 cm = $(67 \times 2.54)/100$ m
$\qquad$ = {$(67 \times 2.54)/100$ m} $\times 10^9$ nm = 1.7×10^9 nm

Example 1.2

How many hydrogen atoms can be lined up end-to-end to make 1 nm length?
$\quad$ Since diameter of a hydrogen atom is 0.1 nm, 1/0.1 or 10 such atoms will be necessary.

1.8 Nanoscience and Nanotechnology

In his visionary 1959 lecture at Caltech, Richard P. Feynman foresaw the potential of the ability to manipulate matter at the atomic scale.

Let us define nanoscience and nanotechnology. *Nanoscience* is the science of matter near or a little above the atomic and molecular level, imbued with special electrical and chemical properties, at dimensions less than 100 nm. *Nanotechnology* is a branch of applied science or engineering concerned with the design, synthesis, characterization, control, manipulation, and application of materials, devices, and systems at the nanoscale (1–100 nm) or having at least one physical dimension of the order of 100 nm or smaller than 100 nm (Booker and Boysen 2008). *How small is 100 nm?* To imagine, it is about one hundred-thousandth the diameter of a human hair (which is 50–100 μm). The size of a bacterium, about the limit of what is visible through most optical microscopes, is about 1 μm (1000 nm). But 100 nm is about the size of a virus, a tenth the size of a bacterium.

Table 1.2 presents the landmarks in the progress of nanotechnology while Table 1.3 traces the evolution of present-day CNT gate field-effect transistor from the triode valve.

Since such materials/devices/systems are found near atomic or molecular scales, nanotechnology deals with objects close to these scales. Further, since properties of matter at nanoscales differ from those in bulk state, nanotechnology is engaged in the study of unusual phenomena and interesting properties of matter that make their appearance because of the small size.

TABLE 1.2

Historical Developments in Nanotechnology

Year	Milestone
1959	Caltech physicist Richard P. Feynman gives his reputed lecture "There is Plenty of Room at the Bottom" on the prospects of atomic engineering
1981	Invention of scanning tunneling microscope by G. Binnig and H. Rohrer of IBM's Zurich Research Laboratory
1985	Discovery of buckyballs by R. F. Curl Jr., H. W. Croto, and R. E. Smalley

TABLE 1.3

Journey from Triode Valve to Carbon Nanotube Transistor

Year	Landmark
1906	Patenting of triode valve (Audion vacuum tube) by Lee De Forest
1947	Invention of transistor by William Shockley, John Bardeen, and Walter Brattain at Bell Laboratories
2003	Filing of patent on carbon nanotube gate field effect transistor by Jeng-hua Wei et al.

In this respect, it involves the extension of natural sciences (physics, chemistry, and biology) into the nanoscale. Nanosciences comprise nanophysics, nanochemistry, nanobiology, nanoelectronics, nanomechanics, nanomachining, and nanomaterials.

1.9 Nanomaterials and the Unusual Behavior at Nanoscales

What type of objects are studied in nanotechnology? The objects under study in nanotechnology are the *nanomaterials*, also called *nanostructured materials* (Bréchignac et al. 2007). As we know, all materials are composed of grains, which in turn are made of molecules and atoms. Nanomaterials are those having grain sizes in the range of nanometers (Ghuo and Tan 2009). These include CNTs, nanowires, nanoscaled thin films, etc. Table 1.4 summarizes the potential applications of nanotechnology. Table 1.5 depicts the drawbacks of nanomaterials.

How are nanomaterials classified? Nanomaterials are divided into three main classes (Table 1.6): (i) *nanoparticles or zero-dimensional (0D) nanomaterials*, for example, atom clusters with particle diameter below 100 nm; (ii) *one-dimensional (1D) nanomaterials* such as nanowires, nanotubes, and nanocables having a width less than 100 nm; and (iii) *two-dimensional (2D) nanomaterials* like nanofilms and superlattices with layer thickness in the nano-range (Cao 2004).

TABLE 1.4

Promises and Expectations from Nanotechnology

Field	Application
Resources	(a) *Energy:* Lighter cars with stronger engines and frames and efficient fuels. House lighting using quantum dots to prevent wastage of energy as heat. Cost-effective solar cells and hydrogen fuel cells. (b) *Water:* Efficient water purification and cheaper desalination plants for supplying water from oceans
Healthcare	(a) *Diagnostics:* Point-of-care lab-on-chip systems for analyzing ailments. New contrast agents floating in the blood stream to light up tumors and for DNA mapping. (b) *Drug delivery:* Reaching the correct amount of medicine precisely at the diseased spot, or nanoshells attaching to cancerous cells and destroying them on laser illumination
Security	(a) *Lightweight but stronger materials* for bridges, aircraft, and skyscrapers. (*b*) *Cheap, disposable but highly sensitive chemical nanosensors* for warning against explosives at airport or anthrax-laced letters. (c) *Quantum cryptography* of data for computers to provide unbreakable security for business and government

TABLE 1.5

Harmful Effects of Nanomaterials

Nanomaterial	Harmful Effect
Buckyballs	Single-walled carbon nanotubes produce lesions in the lungs of rats
Nanoparticles	They could interact with living cells in unanticipated ways. Effects of their inhalation are not known

TABLE 1.6

Classification of Nanomaterials according to Dimensionality

S.No.	Number of Dimensions	Name of the Nanomaterial	Examples
1.	0	Nanoparticle	Buckyball, metallic nanoparticles, e.g., Au, Ag, and Fe
2.	1	Nanowire	Carbon nanotube, silicon nanowire
3.	2	Thin film	Metallic, semiconducting or insulating films, e.g., of Au, Si, and SiO_2

Are nanomaterials found only in the form of elements? No, nanomaterials occur in elemental and also in composite form as compounds or their mixtures and alloys.

What unusual and interesting properties of matter emerge at nanoscales? These properties will be talked about very often in this book. To mention a few interesting properties here, for exemplification purpose, nanomaterials offer greater reactivity, optical absorption, catalytic efficiency, super plasticity (the unusual ability of some metals and alloys to elongate uniformly

by thousands of percent at elevated temperatures) and superparamagnetic characteristics, unique and important aspects of magnetism in nanoparticles. Superparamagnetism is a phenomenon by which magnetic materials may exhibit a behavior similar to paramagnetism even when at temperatures below the Curie or the Néel temperature. The Curie temperature (T_C), or Curie point, is the temperature at which a ferromagnetic or a ferrimagnetic material becomes paramagnetic on heating. The Néel temperature or magnetic ordering temperature, T_N, is the temperature above which an antiferromagnetic material becomes paramagnetic. Superparamagnetism occurs when the material is composed of very small crystallites (1–10 nm). In this case, even when the temperature is below the Curie or Néel temperature (and hence the thermal energy is not sufficient to overcome the coupling forces between neighboring atoms), the thermal energy is able to change the direction of magnetization of the entire crystallite. The resulting fluctuations in the direction of magnetization enable the magnetic field to average to zero so that the material behaves in a manner similar to paramagnetism.

Nanoparticles show higher electrical conductivity, hardness, wear resistance, strength, and toughness. Gold nanoparticles are used as biosensor labels. Pt nanoparticles are superior catalysts. Polymer nanoparticles are used as medicine carriers to target cancerous cells.

Magnetic nanoparticles serve as the new generation of nuclear magnetic resonance (NMR) contrast agents. NMR is the use of magnetic fields and radio waves (instead of the x-rays employed in the CT scan) to visualize the functioning of body structures. A powerful magnet, large enough to enclose the body, forces the atomic nuclei in most material such as soft tissue to align themselves with the magnetic field. Radio waves are aimed at the selected area to excite the atoms, which, on stoppage of the waves, emit signals that are converted to computer-generated pictures. NMR is used in the detection of diseases of soft tissues, heart, brain, and spinal cord.

Zhu et al. (2009) found that the silicon nanowires deformed in a very different way from bulk silicon. Bulk silicon is very brittle and has limited deformability, meaning that it cannot be stretched or warped very much without breaking. But the silicon nanowires are more resilient and can sustain much larger deformation. Young's modulus of Si nanowires was close to the bulk value (187 GPa for <111> orientation; Gigapascal = 103 MPa = 9870 atm = 10^4 bar) when the diameters were larger than 30 nm. As the nanowire diameter decreased, the Young's modulus decreased while the fracture strength increased up to 12.2 GPa. However, the softening trend was obvious when the diameters of the NW were smaller than 30 nm; the Young's modulus of the NWs decreased with the decreasing diameter. Their experiments also showed that fracture strain and strength of Si NWs increased as the NW diameter decreased.

More details about nanomaterials will be given in the ensuing chapters. But the curious reader must be already inquisitive about the reasons for the aforesaid astonishing behavior of matter at the nanoscale. These explanations will be provided as we move through the pages of this book.

1.10 Moving toward Sensors and Transducers: Meaning of Terms "Sensors" and "Transducers"

In our efforts to review basics, let us enquire into the meanings of the terms "sensors" and "transducers" (Wilson 2004). The word "sensor" is derived from the Greek word *sentire* meaning to perceive (Sze 1994). It is a device that converts a physical stimulus such as mechanical motion, heat, light, sound, and magnetic or electric or radiant effect into an electrical signal, which is measured or recorded by an observer or an instrument. It is used for various purposes including measurement or information transfer. For example, a mercury-in-glass (Hg-in-glass) thermometer converts the measured temperature into expansion and contraction of a liquid, which is read on a calibrated glass tube. A thermocouple converts temperature to an output voltage, which is read by a voltmeter.

Similarly, the word "transducer" has originated from the Greek word *transducere*, which means "lead across." It is a device usually electrical, electronic, electromechanical, electromagnetic, photonic, or photovoltaic that converts power from one system to another either in the same or different form. A transducer is a device that is actuated by energy from one system and supplies energy usually in another form to a second system. For example, a motor converts electrical to mechanical energy; hence, it is a transducer. A loudspeaker is a transducer that transforms electrical energy signals into sound energy. An ultrasound transducer transforms electrical signals into ultrasonic waves or vice versa. Similarly, a light emitting diode (LED) converts electric energy into light energy.

What are the differences between a sensor and a transducer? A sensor differs from a transducer in that a transducer converts one form of energy into other form whereas a sensor converts the received signal into electrical form only. A sensor collects information from the real world. A transducer only converts energy from one form to another. The words transducer and sensor are used synonymously or interchangeably by many engineers. Some devices, for example, the microphone, can act as both. The thermostat in home refrigerator takes temperature as the input variable and turns it into mechanical motion by unfolding a bimetallic strip (a strip consisting of two metals of different coefficients of expansion welded together; it buckles on heating) that turns a dial calibrated to read in units of temperature. In this way, the thermal energy associated with a specific temperature

is transduced into mechanical motion. On the contrary, many devices do not play the roles of both the sensor and the transducer. A photoresistor changes resistance as a function of impinging light without any energy conversion. A piezoresistive pressure sensor shows change in piezoresistor value with pressure. The capacitance of a capacitive pressure sensor is altered by pressure.

A plausible way to distinguish between sensors and transducers is to use the term "sensor" for the sensing component itself and "transducer" for the sensing component together with associated circuitry. Transducer = sensor + actuator. Thus, all transducers contain a sensor and mostly, although not always, sensors will also be transducers.

Since all transducers include sensors, therefore, nanotransducers will also include nanosensors. Nanotransducers also come under the purview of this book with an intention to know about their integral sensing parts, that is, nanosensors.

1.11 Definition of Sensor Parameters and Characteristics

Many terms related to sensors will be repeatedly used in the text. It is therefore beneficial to briefly recapitulate the definitions of these terms.

1. *Sensitivity*: It is the change in output value per unit change in input variable (the measurand).
2. *Selectivity*: Ability to detect the measurand in the presence of similar species. It is a measure of the discriminating ability of the sensor with respect to interfering analytes, which are sources of noise in the output.
3. *Resolution*: It is the smallest measurable change in analyte (measurand) value that can be detected by the device.
4. *Response time*: It is the time taken by the sensor to reach $1/e$ or 63% of the final value of the sensed variable (output value).
5. *Calibration characteristic*: It is the curve obtained by plotting the sensor output along the abscissa and the measurand (analyte values) along the ordinate.
6. *Linearity*: It is the degree to which the calibration curve of the sensor matches with a specified straight line approximating the same.
7. *Repeatability*: It is the reproducibility of sensor output readings at given measurand values.
8. *Stability*: It is the degree to which the calibration curve of the sensor remains unchanged over a period of time so that the sensor need not be recalibrated.

9. *Drift*: It is the shift or translation in calibration curve of the sensor with respect to time.

10. *Allowed ambient parameters*: These are permissible maximum values of ambient parameters such as temperature, pressure, relative humidity, light or illumination, etc. under which the device can operate satisfactorily.

1.12 Evolution of Semiconductor-Based Microsensors

Semiconductor-based sensors, mainly fabricated on silicon substrates, are made either in silicon or on silicon by depositing sensitive materials upon it. The earliest semiconductor sensor was a point-contact rectifier in 1904 for detecting radio waves, converting a radiation signal into an electrical one. After the invention of the transistor in 1947, the study of properties of silicon (Si) and germanium (Ge), notably the piezoresistive effect, received a boost. It was found that the resistance changed by two orders of magnitude more than in metallic strain gauges. This led to the use of a silicon bar as a resistive strain gauge by measuring its resistance variation with applied strain. These bars had to be adhesively bonded to metal diaphragms. In subsequent modifications, the piezoresistors were diffused in silicon diaphragms with the diaphragm bonded to a constraint to provide package isolation. In improved versions, the diaphragms were formed by selective anisotropic etching processes, reducing both sensor sizes and costs, and enabling batch fabrication. Wet and dry micromachining techniques were developed, along with silicon wafer-to-wafer and wafer-to-glass bonding methods. Thus, microelectronics was combined with micromachining to fabricate microelectromechanical systems (MEMS)-based sensors in which there were actual moving parts like diaphragms, or membranes, moving bars like cantilevers, etc.

The power requirements of these miniaturized sensors were considerably lower than their traditional counterparts so that signal-processing circuits had to be redesigned for proper interfacing. For long, sensor developments lagged behind tumultuous growth in microelectronics. Therefore, compatibility of the new sensors with circuits, especially on the same chip, was not achieved. A hybrid approach in which the signal-processing circuit was fabricated by thin- and thick-film technologies on a ceramic substrate with active components die mounted and wire-bonded on the substrate was followed.

But the main advantage accruing from silicon microelectronics and MEMS technology–based sensors could be realized by integrating the sensor and the signal-processing electronic circuitry on the same chip. Thus, the semiconductor-based sensors were miniature in size and required

smaller operating power. They offered bulk-manufacturing capability and therefore lower costs when mass-produced. The sizes of these microsensors ranged from hundreds of microns to several millimeters. Note that $1\,\mu m = 10^{-6}$ m.

1.13 From Macrosensor to Microsensor Age and Necessity of Nanoscale Measurements

From the beginning, motivations for sensor miniaturization were clear (Figure 1.8).

1.13.1 A Miniaturized Sensor Can Accomplish Many Tasks That a Bulky Device Cannot Perform

A little contemplation will tell that macro- and microsensors have many limitations. They cannot be used in many situations. Small size and lightweight sensors help in making portable instruments, which are essential for military and aerospace applications besides mobile and handheld consumer products.

To consider an example, a predicted application of nanosensors is in the detection of cancerous cells in the human body by injecting quantum dots. These quantum dots are crystals of semiconductor materials having sizes in the nanometer scale and emitting fluorescent radiation. These crystals are made of cadmium selenide (CdSe), cadmium sulfide (CdS), or indium gallium phosphide (InGaP) and coated with suitable polymers that safeguard
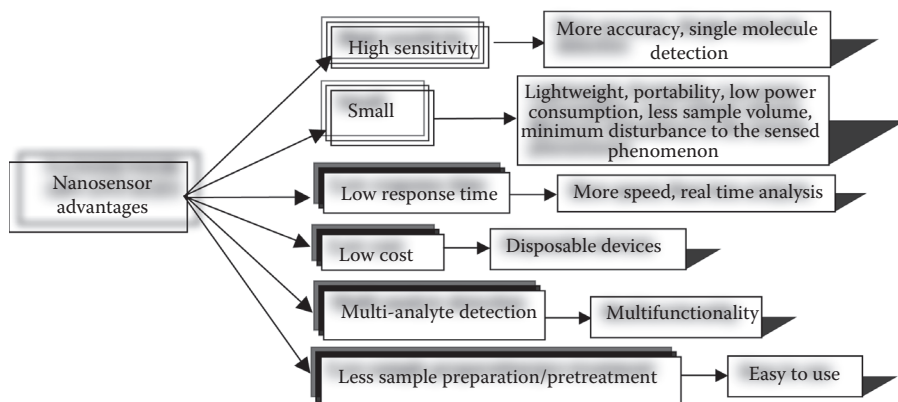
FIGURE 1.8
Advantages of nanosensors.

human cells from toxic action of cadmium and also allow to attach molecules that enable tracking of cell processes and cancers. Obviously, such an application cannot be imagined without developing nanosensors. Macro- or microsensors cannot be inserted in the human body without harming it or disturbing its normal functioning. So it is the small size of nanosensors that is advantageously exploited here. Futuristic advances can open unprecedented perspectives for the application of nanosensors as molecular-level diagnostic and treatment instruments in medicine and as networks of nanorobots for real-time monitoring of physiological parameters of a human body.

System biology, which is currently taking off as research discipline to explore the basic principles of living systems by quantitative modeling of inter- and intracellular processes, will starve for nanosensors and tools to provide data for model verification. Implantable devices like autonomous nanorobots or multifunctional endoscopes (medical devices consisting of a long, thin, flexible or rigid tube that has a light and a video camera, whereby images of the inside of the patient's body can be seen on a screen) for minimal invasive diagnostics, health monitoring, drug delivery, and many other intra-corporal (within a corpus, the body) tasks need ultraminiature sensors to fulfill their missions while minimizing invasiveness. As limitations in downscaling of conventional systems are foreseeable, new materials with new properties on the nanoscale will emerge to fulfill sensor tasks in ultraminiaturized sensor systems.

1.13.2 The Issue of Power Consumption

Conventional sensors were bulky and power-hungry. For example, consider a metal oxide gas sensor. An earlier version consists of a ceramic (hard, brittle, heat-resistant, and corrosion-resistant material) tube with a heater coil fixed axially inside. A paste made from (tin oxide (SnO_2) powder + palladium (Pd)) was applied over the outer surface of the ceramic tube. The heater raised the temperature of the tin oxide layer to 200°C–500°C at which the sensitivity to inflammable gases increases. The sensing principle was the decrease in resistance of tin oxide in the presence of a combustible gas in the ambient. Obviously, the sensor consumed several tens of watts of power during operation. The next generation of sensors consisted of a ceramic substrate. On the bottom side of the substrate, a meander-shaped Pt heater was deposited. On the topside, the tin oxide paste was applied by screen printing (a printing technique that uses a woven mesh to support an ink-blocking stencil; a roller or squeegee is moved across the screen stencil, forcing ink past the threads of the woven mesh in the open areas). This sensor had lower power consumption than its predecessor, a few watts.

It is evident that gas sensing requires neither a large volume of material (only the surface area must be large) nor a thick layer of sensing material. Therefore, most of the power used in heating is unnecessarily wasted. Considerable power saving could be achieved by reducing the thermal mass

of the sensor. So, the next improvement consisted in fabricating a miniaturized hotplate by micromachining of silicon. This hotplate fabricated by the MEMS technology was as small as $100\,\mu m \times 100\,\mu m$. The thickness of the supporting plate (silicon dioxide + silicon nitride) was $2\,\mu m$, and the hotplate was suspended over a cavity to minimize conduction losses. Such a hotplate and, therefore, the resulting sensor hardly consumed milliwatts of power. However, reduction in power consumption is only one of the reasons for downsizing sensors.

1.13.3 Low Response Times

Small-size sensors require less time to reach equilibrium with the sensed environment because the signals take less time to traverse shorter path lengths. Therefore, these sensors can implement the desired operations at faster speeds.

What benefit is derived from a fast speed? Measurements are performed in real time and data are immediately analyzed for necessary action.

1.13.4 Multi-Analyte Detection and Multifunctionality

If one has to detect several gases, an array of gas sensors will be required. To construct such an array consisting of hundreds or thousands of sensors will lead to an enormously bulky and costly sensor, which will be virtually impractical to fabricate and use. The multiplexing capability offered by nanosensors is a vast improvement for real-time gas composition monitoring. An array of thousands of nanosensors, each coated with a different functional group and hence tuned to a prefixed analyte, can be used in a single device. Such an ultraminiaturized, low-power device supplemented with signal-processing and pattern-recognition algorithms will exhibit effective discrimination capabilities among target analytes, producing a unique signature or fingerprint on exposure to a specimen containing a mixture of chemical and biological characters required for environmental pollution control.

The above situation is comparable with the historical development of computers. The primitive computers were so big that they required full rooms for their storage, using high power and producing heat. With advancement of technology, the size decreased reaching desktop personal computers, then laptops, and then palmtop computers. Power consumption was drastically reduced in this progression reaching very low levels in battery-operated computers. At the same time, the capabilities of the computers also improved. This was enabled by fast strides made by very-large-scale and ultra-large-scale ICs, Very-Large-Scale Integration (VLSI), and Ultra-Large-Scale Integration (ULSI).

Multifunctionality is a versatile feature in nature. The human tongue is a taste sensor and also used in speech. Human nose is used for

smelling and breathing. Similarly many other organs serve multiple functions. Multifunctionality is also expected from nanosensors.

1.13.5 Sensitivity Considerations and Need of Functionalization

Many sensors are based upon the adsorption of target analytes and adsorption depends on surface area and surface chemistry. Making a sensor out of a nanoporous or nanocrystalline material increases the surface area tremendously, for example, the surface area of a single-walled carbon nanotube (SWCNT) is $1600\,m^2\,g^{-1}$. Sensitivity of a gas sensor is augmented several-fold by increasing the surface area. By using tin oxide (SnO_2), indium oxide (In_2O_3), antimony oxide (Sb_2O_3) or zinc oxide (ZnO) in nanoparticle form, the sensitivity is considerably enhanced.

On their own, the CNTs have poor sensitivity toward analytes but by surface modification such as by suitably coating the CNTs or by doping palladium (Pd) atoms, they are made selective toward particular species.

Many types of microelectronic platforms are made. Notable examples are microhotplate, ion-sensitive field-effect transistor (ISFET) and microcantilever. The ISFET is a MOSFET without the metal layer on the gate and with gate dielectric made of materials like Si_3N_4, Al_2O_3, and Ta_2O_5. It is a pH sensor that is converted into a specific ion sensor or a biosensor by coating suitable membranes on the gate.

The above platforms are inherently not sensitive to any analyte or sensitive to specific analytes only, and therefore need to be coated with suitable materials to make them sensitive thereby enabling the development of families of sensors. Thus, each platform is a source of a sensor family.

1.13.6 Interfacing with Biomolecules

Nanomaterials have sizes comparable with those of biomolecules like proteins, virus (a small infectious agent), cells, nucleic acids (the building blocks of living organisms), etc., and thus can form easy interfaces between biomolecules and readout instruments. Equality or comparability of sizes of nanomaterials and biomolecules is exploited in nanosensors. Many nanosensors are based on the interfacing of nanomaterials with biomolecules. These are known as nanobiosensors.

1.13.7 Low Costs

Sensor miniaturization and device integration, based on reproducible fabrication processes and large-scale production, are the top prerequisites for low-cost products, and these requirements are fulfilled by nanosensors so that research on nanosensors will lead to cheaper devices. Availability of disposable nanosensors in abundant supply at throw-away prices is the need of the hour.

1.13.8 Possibility of a New Genre of Devices

Nanosensors allow for building an entirely new class of devices that provide the elemental base for "intelligent sensors" capable of data processing, storage, and analysis. These sensors will provide high accuracy, ultrahigh sensitivity, extreme specificity, real-time *in vivo* information with greater speed, having multi-analyte options, requiring smaller quantity of sample and minimal sample preparation, durable, safe, and portable. Some members of the nanosensor generation of devices have already qualified in laboratory tests, and are starting to appear in the marketplace. But there remains a long way to tread and several issues to be solved. Nonetheless, the potential is enormous.

1.14 Definition and Classification of Nanosensors

Definition of nanosensors must be clear. In fact, any sensor that uses a nanoscale phenomenon for its operation is a nanosensor. Any sensor fabricated by nanotechnological methods is a nanosensor, that is, nanosensor is a nanotechnology-based sensor. To define the term, let us agree that any sensor characterized by one of the following properties will be labeled as a nanosensor: (i) the size of the sensor is in nanoscale, (ii) sensitivity of the sensor is in the nanoscale, or (iii) the spatial interaction distance between the sensor and the object is in nanometers. In this book, whenever a nanosensor is talked about, the implication is that at least one of these criteria is valid.

From definition (i), nanosensors involve signal transformation from the environment using nanostructures, that is, structures having at least one dimension in the lateral direction less than 100 nm and the other less than 1 μm. Hence, in this definition, sensors must comply with nanotechnology based on geometrical dimensions of the sensor (Figure 1.9).

A nanowire is a wire of a material, for example, a metal, the diameter of which is less than 100 nm. Nanofibers are submicron-sized fibers whose diameter is 50–500 nm. Nanotubes are hollow cylinders only a few nanometers wide, made of one element such as carbon. Nanobelts are nanostructures in the form of belts. Nanoprobes are optical devices for viewing extremely small objects. Quantum dots are nanosized fluorescent semiconductor crystals. Thin films of interest here are those having thickness in nanometer range.

Many conventional techniques for nanometer-precision measurements or position control do not fulfill the criterion of definition (i). But they are in conformance with the criteria of definitions (ii) and (iii). Many authors exclude them from the domain of nanosensors. But in this book all such sensors that are not necessarily small in size, which may be having micro- or
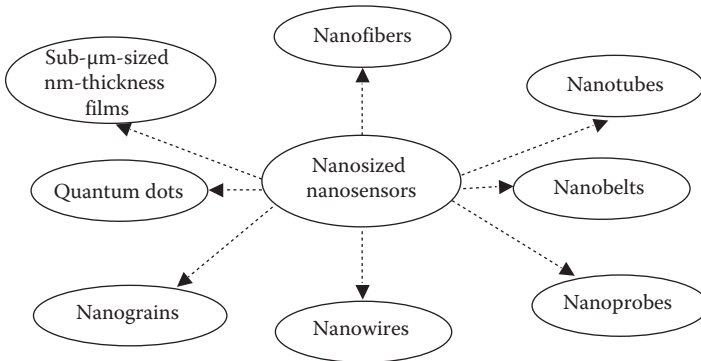
FIGURE 1.9
The nanosized nanosensor family tree.

macro-dimensions, will be categorized as nanosensors if their sensitivities fall in nano-units, for example, nm for position or displacement sensor, nN for force sensor, nK for temperature sensor, nM for solution concentration sensor, and so on. Moreover all such sensors in which spatial interaction distance between the investigating electrode and the examined object is a few nanometers, such as in atomic force microscope (AFM), are also nanosensors (Figure 1.10). Thus, in this book, interpretation of the term "nanosensors" is broad and this will provide a wide perspective of the field, encompassing a wide diversity of nanosensors in the scope and coverage of the book.

Based on earlier discussions, nanosensors can be broadly classified into two groups, namely, those that have nanoscale dimensions and those that perform nanoscale measurements but do not necessarily have nanoscale dimensions. In the former group are two subgroups: sensors having all the three dimensions in nanometers like quantum dots (fluorescent nanosized
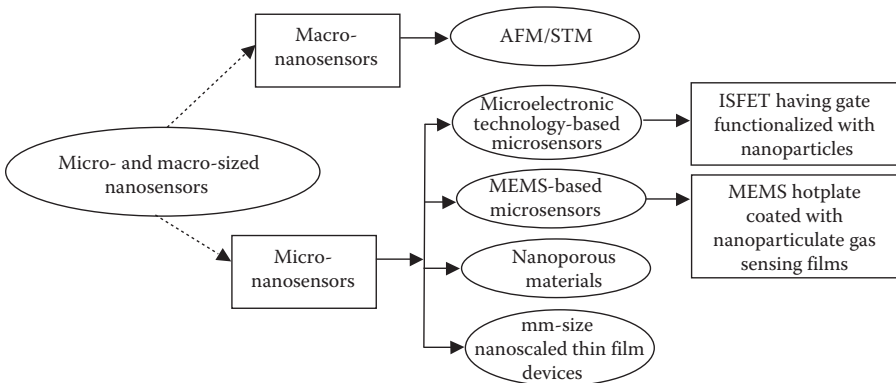
FIGURE 1.10
Micro- and macro-sized nanosensors.

TABLE 1.7

Conversion Phenomena of Sensors

Stimuli	Phenomena
Physical	Thermoelectric, photoelectric, photomagnetic, magnetoelectric, electromagnetic, thermoelastic, elastoelectric, thermomagnetic, thermo-optic, photoelastic
Chemical	Chemical transformation, physical transformation, electrochemical process, spectroscopy
Biological	Biological transformation, physical transformation, effects on test organisms, spectroscopy

semiconductor crystals) and nanoscale thin films like 100–500 Å thick films of metal oxides, palladium, etc., used for gas sensing, which have only one nanoscale dimension.

Further, like all sensors, nanosensors can be classified as active (those that require an energy source, for example, a thermistor [thermally sensitive resistor], a resistive strain gauge, etc.) and passive (those that do not require the energy source, for example, a thermocouple, a photodiode [a P–N diode in which the reverse leakage current changes with illumination], a piezoelectric sensor [a device that uses the piezoelectric effect to measure pressure, acceleration, strain, or force], etc.) nanosensors.

Another classification scheme consists of two categories: absolute and relative nanosensors, for example, an absolute pressure sensor measures pressure with reference to zero pressure (vacuum) while a relative pressure sensor does the same with respect to an arbitrarily chosen reference such as atmospheric pressure. An absolute sensor detects a stimulus with respect to an absolute physical scale independent of measurement conditions while the relative sensor produces a signal in reference to some special chosen case.

Proceeding further, nanosensors (like sensors in general) are classified according to the form of energy signal detected such as physical (mechanical, thermal, optical) chemical, and biological (Table 1.7).

Advancing still further, sensor classification is based on the measurand like a mechanical variable, for example, position, displacement, or a chemical variable like concentration of a sample (Table 1.8).

1.15 Physical, Chemical, and Biological Nanosensors

Let us elaborate on the sensor classes: physical, chemical, and biological. *Physical sensors* are used for measuring properties like temperature, pressure, flow, stress, strain, position, displacement, or force. *Chemical sensors* are meant for determining concentration or identity of a chemical substance like

TABLE 1.8

Stimuli and Measurands

Stimulus	Properties
Mechanical	Position, acceleration, stress, strain, force, pressure, mass, density, viscosity, moment, torque
Acoustic	Wave amplitude, phase, polarization, velocity
Optical	Absorbance, reflectance, fluorescence, luminescence, refractive index, light scattering
Thermal	Temperature, flux, thermal conductivity, specific heat
Electrical	Charge, current, potential, dielectric constant, conductivity
Magnetic	Magnetic field, flux, permeability
Chemical	Components (identities, concentrations, states)
Biological	Biomass (identities, concentrations, states)

ethanol (C_2H_5OH), carbon monoxide (CO), gasoline or petrol, or other molecules. Biosensors are useful for dealing with biologically active substances whether they are cellular like toxic plague bacteria (*yersinia pestis* found mainly in rodents, particularly rats) or anthrax spores (*bacillus anthracis* bacterium causes anthrax; it forms dormant endospores, tough structures, somewhat like bacterial seeds, allowing bacteria to go dormant under unfavorable conditions, often referred to as spores, and multiplying rapidly upon inhalation, ingestion, or coming into contact with a skin lesion on a host), supramolecular (collective behavior of organized ensembles of molecules) like flu (influenza) viruses, or molecular like protein toxins such as Staphylococcal enterotoxin B (SEB), an enterotoxin excreted by the *Staphylococcus aureus bacterium*. Sometimes biosensors are considered as a subset of chemical sensors. In this book, this classification of nanosensors will be followed. The measurands for physical, chemical, and biological sensors are as follows:

1. *Physical*
   a. *Mechanical*: mass, density, viscosity, position, velocity, flow rate, acceleration, force, stress, strain, moment, and torque
   b. *Acoustic*: wave velocity, amplitude, phase, polarization, and spectrum
   c. *Thermal*: temperature, flux, thermal conductivity, and specific heat
   d. *Radiation*: type, energy, and intensity
   e. *Optical*: wave velocity, amplitude, phase, and polarization
   f. *Magnetic*: magnetic permeability, field, and flux
   g. *Electric charge*: current, potential, potential difference, permittivity, capacitance, and resistance

2. *Chemical*

Identities, concentrations, and states of elements or compounds

3. *Biological*

Identities, concentrations, and states of biomolecules

1.16 Some Examples of Nanosensors

1.16.1 Common Nanosensors

At present, the most commonly observed nanosensors exist in nature in the biological world as natural receptors of outside stimulation. Animals like dogs have a strong sense of smell that functions using receptors that feel nanosized molecules. Various fishes use nanosensors to perceive minuscule vibrations in the surrounding water; several insects detect sex pheromones using nanosensors. Similar to animals, many plants also use nanosensors to detect sunlight. In the artificial world, most film cameras have used photosensors at the nanosize for years. Traditional photographic film uses a layer of silver ions that become excited by solar energy and cluster into groups, as small as four atoms apiece in some cases, scattering light and appearing dark on the frame. Various other types of films can be made using a similar process to detect other specific wavelengths such as IR, ultraviolet, and x-rays.

1.16.2 Carbon Nanotube–Based Nanosensors

Because of the ability to functionalize CNTs to detect specific molecules and the changes in their dielectric constants and permittivities with adsorption of gases or vapors, a family of nanotube-based sensors has emerged.

1.16.3 Nanoscaled Thin Film Sensors

Nanoscaled thin films of polymers and certain oxides have allowed pre-concentration of analytes, for example, heavy metals (metals with a specific gravity greater than about 5.0, especially those that are poisonous: antimony (Sb), arsenic (As), bismuth (Bi), cadmium (Cd), cerium (Ce), chromium (Cr), cobalt (Co), copper (Cu), gallium (Ga), gold (Au), iron (Fe), lead (Pb), manganese (Mn), mercury (Hg), nickel (Ni), platinum (Pt), silver (Ag), tellurium (Te), thallium (Tl), tin (Sn), uranium (U), vanadium (V), and zinc (Zn)) by functionalized nanoporous thin films for x-ray fluorescence analysis, extending the limits of detection from ppm to ppb level.

1.16.4 Microcantilever- and Nanocantilever-Enabled Nanosensors

Arrays of suitably functionalized cantilevers are used to detect explosives as lower-cost alternatives to mass spectrometers; the mass spectrometer is an analytical instrument that measures the mass-to-charge ratio of charged particles.

1.17 Getting Familiar with Analytical and Characterization Tools: Microscopic Techniques to View Nanomaterials and Nanosensors

Nanomaterials and many features of nanosensors are invisible to the naked eye. But several instruments have helped in viewing and understanding these devices.

As already mentioned in Section 1.3.8, light is an electromagnetic wave comprising electric and magnetic fields representing energy, not matter; hence, it can travel through vacuum. High-frequency light has more energy (in familiar water analogy, more waves strike the sea shore per second delivering greater energy). Light is not visible below red (IR) or beyond violet (ultraviolet). IR light has lowest frequency and energy while ultraviolet light has highest frequency and energy. The continuum of light frequencies and energies constitutes its electromagnetic spectrum.

Does light behave as waves only? Light also behaves in some ways as if it consists of discrete particles. These apparent particles or energy packets have been designated as *photons*. According to wave–particle duality concept, objects show wave- or particle-like behavior depending on the experiment being performed. When an object is heated, it becomes red hot and then white hot. Electrons are excited from normal (ground state) orbitals to excited states. Falling back to the ground states, they emit photons or light.

What is a microscope, and what is meant by microscopy? Microscope is an instrument that uses an optical lens or a combination of lenses, electronic or other processes, to produce magnified images of small objects, specially those objects that are too small to be seen by the unaided eye. A microscope augments the power of the eye to see small objects. Microscopy is the technical field of using microscopes to view samples or objects. It deals with the examination of minute objects by means of a microscope, and is the science of interpretive use and application of microscopes for research.

Which is the most common microscope? Is it useful for nanotechnology? The most familiar and widely used microscope is the optical or light microscope but it is of little use in nanotechnology. Microscopes for nano-regime exploration will be described later with respect to their operating principles, applications, and limitations. Use of these microscopes requires special attention to

sample preparation. Some microscopes can probe conducting samples only; others can probe both conducting and insulating specimens. Some can work in vacuum only; others can operate in air or even under liquids. Therefore, the operational and other salient features of microscopes are presented.

1.17.1 Scanning Electron Microscope

How does a scanning electron microscope (SEM) produce the picture of the surface of an object? In an SEM, a stream of electrons called the *primary electrons* bombards objects under examination in vacuum to create their images (Goldstein et al. 2003). These primary electrons dislodge electrons from the sample, and the emitted secondary electrons are collected by a positively charged grid (Figure 1.11). As the primary electron beam scans the surface of the sample, secondary electrons are obtained from different portions of the sample. The secondary electrons thus gathered are used to construct the images of small invisible objects on the scale of 10 nm on a computer monitor.

What information is provided by SEM pictures? SEM images provide information on the topography of the surface such as its texture and morphology including the shape, size, and positions of particles on the surface of the sample.
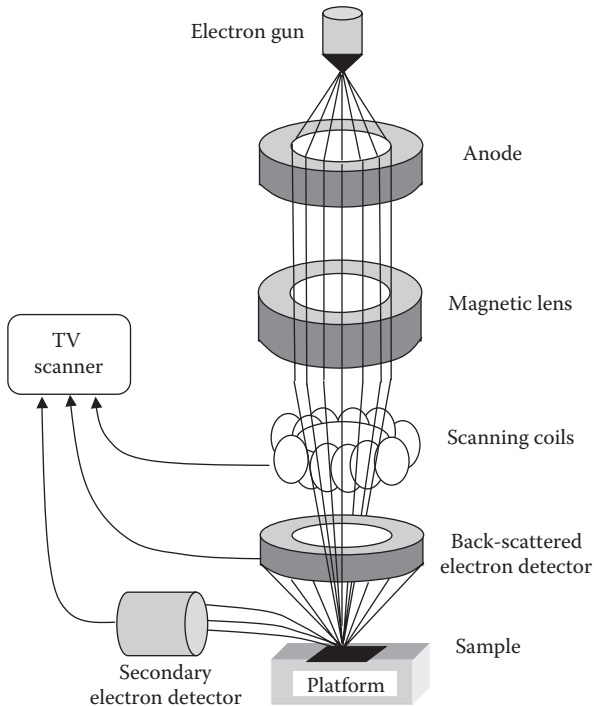
FIGURE 1.11
Image construction process by SEM (http://www.purdue.edu/rem/rs/graphics/sem2.gif).

What are the limitations of SEM? SEM is restricted to surface studies only, and is not useful for exploring the interior of samples. *What will happen if the sample is an insulator?* The electron cloud produced on the sample surface will be stored there. This is called charging effect and hinders the imaging of insulating surfaces. For discharging the surface, a thin conducting gold film having a thickness of a few hundred angstroms is deposited on the sample surface. This helps in imaging an insulating sample by removing the charge buildup by electron cloud.

1.17.2 Transmission Electron Microscope

How does transmission electron microscope (TEM) differ from SEM in producing an image? In contrast to SEM, which scans a surface to produce its image, in a TEM, the electron beam is transmitted through a sample to observe the density of its constituents and thus produce an image through density variations. SEM works by surface bombardment with electrons and secondary electron collection. In TEM, electrons penetrate and cross to the opposite side of the sample.

In TEM, a beam of electrons passes through a sample to produce a projected image on a phosphor screen. A large number of electrons are able to pass through less dense regions of the sample. These are observed as light areas. On the other hand, a smaller number of electrons are able to cross the denser regions of the sample giving darker areas in the image. Thus, an *inverse density image* of the sample is obtained.

How do SEMS and TEMS differ in terms of resolution, sample preparation, costs, and applications? TEMs give greater image resolution (the details an image holds) than SEMs but TEM sample preparation procedure is extremely difficult. Also, TEMs are more expensive instruments. Therefore, SEMs are used as workhorses for routine work whereas TEMs are engaged in specialized tasks.

1.17.3 Scanning Tunneling Microscope

Unlike Newtonian mechanics, wave mechanics allows the flow of current called tunneling current across a thin barrier. If the surface to be probed is conducting, the tunneling current flow between a conducting tip and surface of the sample at small separations of few nanometers between them helps in constructing images of the surface (Figure 1.12). This is the principle of the scanning tunneling microscope (STM).

How to examine insulating specimens? Obviously, other microscopes like AFM must be resorted to.

1.17.4 Atomic Force Microscope

How does an AFM produce an image? In the AFM, the tip of a cantilever moves on the surface of the sample. Interatomic van der Waals forces (weak electric
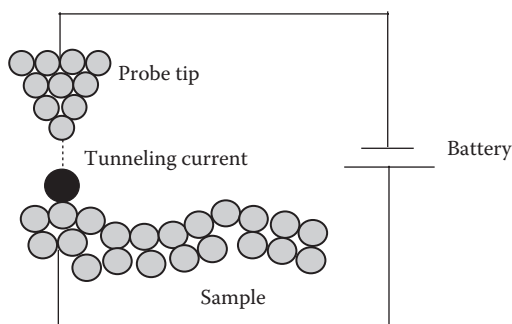
FIGURE 1.12
Principle of STM.

forces that attract neutral molecules to one another caused by a temporary change in dipole moment arising from a brief shift of orbital electrons to one side of one atom or molecule, creating a similar shift in adjacent atoms or molecules) act between the tip and the sample (Figure 1.13). As the tip moves across the hills and valleys on the surface, a laser beam is reflected from the back of the cantilever into an array of photodetectors. This reflected laser beam generates a profile of the surface of the sample (Figure 1.14).

What are the distinguishing features of AFM in relation to other microscopes? Unlike SEM, AFM can operate in air and even underneath liquids. This is obviously a great advantage. Vacuum is used in AFM to prevent contamination of the sample, if necessary. Also, AFM produces a three-dimensional (3D) image of the sample in contrast to 2D images obtained by SEM. This must be carefully noted. Further, AFM is used to study conducting, semiconducting, or insulating samples. As previously remarked, in SEM, a thin conducting film must be deposited over an insulating sample to prevent charge buildup on its surface. In an STM, the sample must be conducting.
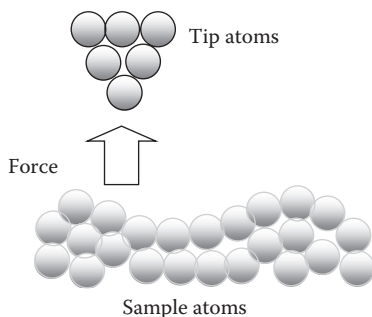
FIGURE 1.13
Operating principle of AFM.
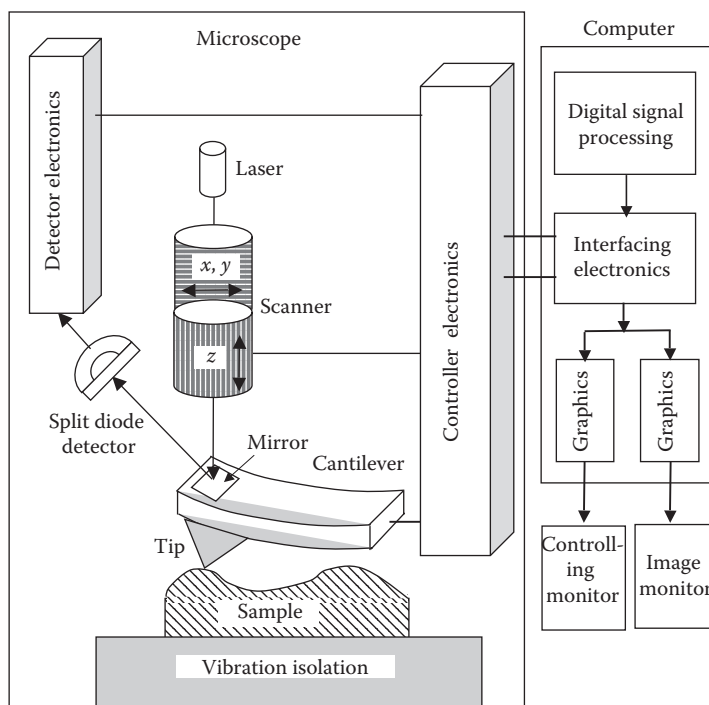
FIGURE 1.14
Parts of AFM (http://nano.tm.agilent.com/index.cgi?CONTENT_ID=809; http://nano.nd.edu/SC190/scanningprobe.pdf).

How does the tip of AFM affect its resolution? The sharper the AFM tip, the better is the resolution. A CNT serves as a good material for AFM tip. As it is cylindrical in shape, it provides a more accurate profile of the sample surface because it can penetrate into deep crevices on the surface where a pyramidal-shaped tip cannot reach; a pyramid is a structure where the outer surfaces are triangular and converge at a point. Moreover, it is more flexible and stronger than common silicon or silicon nitride tips.

Table 1.9 displays the characteristic features of various microscopes in a nutshell. The reader must take notice of and remember these features.

1.18 Spectroscopic Techniques for Analyzing Chemical Composition of Nanomaterials and Nanosensors

What is a spectrum? A spectrum is the distribution of a characteristic of a physical system or phenomenon, such as the distribution of energy or mass, measured as intensity, emitted by a radiant source in order of wavelengths,

TABLE 1.9

A Look at the Features and Capabilities of Different Microscopes

	Optical	SEM	STM	AFM
Imaging environment	Air, liquid, vacuum	Vacuum	Air, liquid, vacuum	Air, liquid, vacuum
Sample preparation requirement	Easy	Little to considerable	Easy	Easy
Sample restrictions	Must not be transparent to wavelength of light used	Charge buildup on the sample is to be avoided	(i) Conducting only (ii) Height variations in the sample should be <10 μm	Height variations in the sample should be <10 μm
In situ imaging possibility	Yes	No	Yes	Yes
In-fluid imaging possibility	Yes	No	Yes	Yes
Depth of field	Small	Large	Medium	Medium
Depth of focus	Medium	Large	Small	Small
Resolution (x, y)	1.0 μm	5 nm	0.1 nm	2–10 nm
Resolution z	Not applicable	Not applicable	0.05 nm	0.05 nm
Effective magnification	$1–2 \times 10^3$	$10–10^6$	$5 \times 10^2–10^8$	$5 \times 10^2–10^8$

frequency, or some other property. Examples are electromagnetic spectrum, visible spectrum, and ultraviolet spectrum. A mass spectrum is represented as a vertical bar graph, in which each bar represents an ion having a specific mass-to-charge ratio and the length of the bar indicates the relative abundance of the ion.

What is meant by spectroscopy? Spectroscopy is concerned with identifying elements and compounds and elucidating atomic and molecular structure by measuring the radiant energy absorbed or emitted by a substance at characteristic wavelengths of the spectrum because such emission or absorption behavior is a unique property useful for recognizing presence of the element or compound.

What does a spectrometer do? A spectrometer is used for producing spectral lines and measuring their intensities and wavelengths (Perkampus and Charlotte Grinter 1992).

1.18.1 Infrared Spectroscopy

What are IR rays? IR radiation consists of electromagnetic waves. It occupies a part of the electromagnetic spectrum having wave numbers from 13,000 to 10 cm^{-1}, or wavelengths between 0.78 and 1000 μm; wave number is the reciprocal of wavelength.

What is IR spectroscopy? It is a type of vibrational spectroscopy concerning the absorption measurement of different IR frequencies by a sample placed in the path of an IR beam (Hollas and Hollas 2004).

What is the principle of IR spectroscopy? The basis of IR spectroscopy is that IR radiation is in the same frequency range as a vibrating molecule. This is essentially the reason for choosing IR radiation for this purpose. Therefore, when a vibrating molecule is bombarded with IR rays, it absorbs those frequencies in the incident rays, which exactly match the frequencies of the different harmonic oscillators making up that molecule. When this radiation is absorbed, the little oscillators in the molecule continue vibrating at the same frequency, but since they have absorbed the energy of the light, they have larger amplitude of vibration. The remaining unabsorbed light by any of the oscillators in the molecule is transmitted through the sample to a detector.

IR absorption information is generally presented as a spectrum in which wavelength or wave number (the number of complete wave cycles of a wave that exist in one meter [1 m] of linear space; it is expressed in reciprocal meters [m^{-1}]) is plotted as the abcissa and absorption intensity (the ratio of the radiant flux absorbed by a body to that incident upon it) or percent transmittance (% incident light at a specified wavelength that passes through a sample) as the ordinate. A computer analyzes this spectrum and determines the absorbed frequencies. Using Fourier transform (a mathematical operation that decomposes a signal into its constituent frequencies) algorithm (a set of instructions for solving a problem), one can bombard the molecule with every frequency of IR radiation at one time, and obtain a spectrum (Smith 1995).

How are the compounds identified by IR spectroscopy? IR spectrum of a given compound is unique and therefore serves as a *fingerprint* for this compound. IR spectroscopy is useful for identifying certain functional groups (collections of atoms in a molecule that participate in characteristic reactions) in molecules. But identification of an unknown compound by IR spectroscopy alone is rarely possible.

What are the applications and limitations of IR spectroscopy? It can identify polar molecules (molecules that have mostly positive charge on one side and mostly negative charge on the other; this charge difference allows the positive end of a molecule to attract to the negative end of another) but is not useful for nonpolar molecules (molecules that share electrons equally and do not have oppositely charged ends). It is utilized for recognizing certain functional groups in molecules but it is rarely, if ever, possible to identify an unknown compound by using IR spectroscopy alone. It is a handy tool for studying SWCNTs.

1.18.2 Ultraviolet–Visible Spectroscopy

Do molecules absorb only IR radiation? No, molecules absorb visible and even UV radiation (UV = 200–400 nm, visible = 400–800 nm). Ultraviolet–visible

(UV–vis) is spectroscopy that employs UV or visible light source for irradiating the sample. Electrons in different atoms absorb certain frequencies of light, reaching an "excited state." The excited atom gives off a photon to return to its original or ground state.

How does the principle of UV–vis spectroscopy differ from IR spectroscopy? Unlike IR spectroscopy (which views vibrational motions), UV–vis spectroscopy looks at electronic transitions. Applying Beer's law, $A = \varepsilon bc$ (where A = absorbance, ε = molar extinction coefficient, b = path length, and c = concentration), and from the absorbance information, one is able to determine either the concentration of a sample if the molar extinction coefficient (a property of chemical species that measures the strength with which it absorbs light at a particular wavelength per molar concentration) is known, or the molar absorptivity, from knowledge of concentration.

What property helps in identification of compounds? As molar extinction coefficients are specific to particular compounds, UV–vis spectroscopy is helpful in determining the identity of an unknown compound. The spectrometer measures the frequencies of radiation passing through the sample and the missing frequencies reveal the identities of atoms and molecules present in the sample.

For what purpose is UV–vis spectroscopy especially useful? UV–vis spectroscopy is commonly used to study samples that IR spectroscopy cannot identify.

1.18.3 Raman Spectroscopy

The reader must have noted that starting with using vibration motions of atoms and molecules in IR spectroscopy, there is a progression toward applying electronic transitions and photon emissions in UV–vis spectroscopy. Raman spectroscopy is related to photon emission in Raman effect.

How does Raman spectroscopy utilize Raman effect and what is this effect? Raman spectroscopy is a spectroscopic technique based on *inelastic scattering of monochromatic light* from a laser source (Ferraro 2003). In inelastic scattering, the frequency of photons in monochromatic light (consisting of radiation of a single wavelength or of a very small range of wavelengths) changes upon interaction with a sample. Photons of the laser beam are absorbed by the sample and then reemitted. Frequency of the reemitted photons is shifted toward higher (Stokes scattering) or lower values (anti-Stokes scattering) in comparison with original monochromatic frequency, which is called the *Raman effect*. There are two types of Raman scattering: Stokes scattering and anti-Stokes scattering. In Stokes scattering, the electron decays to a state higher than its original state while in anti-Stokes scattering, the electron decays to a state lower than the original state.

Reradiation of energy at the same frequency as the original light is known as *elastic Rayleigh scattering*. Stokes and anti-Stokes scattering spectra are mirror images on opposite sides of the Rayleigh line. Stokes

scattering spectra are less temperature sensitive; hence, they are generally used.

Raman spectroscopy is applied to study solid, liquid, and gaseous samples. Unlike IR spectroscopy, Raman spectroscopy is capable of identifying nonpolar molecules in which the electrons are shared equally between the nuclei. Their charge distribution is spherically symmetric when averaged over time. Consequent to this even distribution of charge, the force of attraction between different molecules is small.

What kind of irradiation source is used in Raman spectroscopy? The Raman effect is very weak. To obtain meaningful results, an intense laser source is necessary for irradiating the sample.

1.18.4 Energy Dispersive X-Ray Spectroscopy

It is a technique based on the use of x-rays generated during the SEM examination of a sample. The x-rays produced on irradiating the sample (10–20 keV) in SEM give information about the chemical composition of the sample (Goldstein et al. 2003). In this method called *energy dispersive x-ray (EDX) spectroscopy*, a detector collects x-rays and produces current pulses proportional to energies of x-rays, enabling the acquisition of an image of each element in the sample.

How are these current pulses used to recognize elements? There are peaks in x-ray energies corresponding to different elements. By reading the positions of energies of these peaks for a given sample, one is able to identify the elements present in it by comparing the energies of peaks obtained in the sample under study with the chart of samples of known compositions and peak positions.

1.18.5 Auger Electron Spectroscopy

Electrons of energy 3–20 keV bombard a sample (Briggs and Grant 2006). A transference of energy occurs, exciting a core electron into an orbital of higher energy. On reaching this excited state, the atom has two possible modes of relaxation: emission of an x-ray or an Auger electron. In x-ray emission, the energy is given off as a single x-ray photon. In the Auger process, the electron transfers its energy to an inner orbital electron and this latter electron, called the "Auger electron," is ejected. It carries a characteristic kinetic energy.

How do the processes described earlier help in recognition of species? In both the relaxation processes, the emitted particle has an energy peculiarity (trait) of the parent element. An energy spectrum of the detected electrons therefore shows peaks assignable to the elements present in the sample.

What are the uses and limitations of the Auger electron spectroscopy (AES) method? AES is a popular technique for determining the composition of the top few layers of a surface. It is unable to detect hydrogen or helium, but is sensitive to all other elements. It is most sensitive to the low atomic number elements.

1.18.6 X-Ray Diffraction

X-ray powder diffraction (XRD) is a rapid analytical technique primarily used for phase identification of a crystalline material. A diffraction pattern is obtained when x-rays interact with a crystalline substance (Phase). The diffraction pattern is the distinctive pattern of light and dark fringes, rings, etc., formed by diffraction phenomenon. It is an interference pattern produced when a wave or a series of waves undergoes diffraction, as when passed through a diffraction grating or the lattices of a crystal. This pattern provides information about the frequency of the wave and the structure of the material causing the diffraction, that is, the x-ray diffraction pattern of a pure substance is like a fingerprint of the substance. The powder diffraction technique identifies components in a sample by a search/match procedure.

1.18.7 X-Ray Photoelectron Spectroscopy or Electron Spectroscopy for Chemical Analysis

In x-ray photoelectron spectroscopy (XPS), a high-energy x-ray source (*hv*) is used for the excitation and ejection of core level electrons. Al Kalpha (1486.6 eV) or Mg Kalpha (1253.6 eV) are frequently chosen as the photon energies. The energy of the photoelectrons (electrons released or ejected from a substance by photoelectric effect, that is, under irradiation with photons of sufficiently low wavelength) leaving the sample is determined giving a spectrum with a series of photoelectron peaks. The binding energy of the peaks is characteristic of each element. XPS is insensitive to hydrogen (H) or helium (He), but detects all other elements.

1.18.8 Secondary Ion Mass Spectrometry

The surface of the sample is bombarded with high-energy ions. As a result, ions and atoms are sputtered from the surface. Sputtering is the process of ejection of ions from the surface by the impact of the bombarding ions. These sputtered ions are detected and mass analyzed to determine the elements from which the sample is made. Thus, bombardment of a sample surface with a primary ion beam followed by mass spectrometry of the emitted secondary ions constitutes secondary ion mass spectrometry (SIMS).

In order to measure the characteristics of individual molecules, a mass spectrometer converts them to ions so that they can be moved about and manipulated by external electric and magnetic fields. Ion formation and manipulation is conducted in high vacuum. By changing the strength of the magnetic field, ions of different masses are focused on a detector fixed at the end of a curved tube under a high vacuum (10^{-5}–10^{-8} torr). A mass spectrum is a vertical bar graph. In this bar graph, each bar represents an ion having a specific mass-to-charge ratio (*m*/Z) and the length of the bar indicates the relative abundance of the ion.

Mass spectrometers can easily distinguish ions differing by only a single amu. Thus, they provide accurate values for the molecular mass of a compound. The highest-mass ion in a spectrum is considered to be the molecular ion. Lower-mass ions are fragments from the molecular ion, assuming that the sample is a single pure compound.

1.19 The Displacement Nanosensor: STM

1.19.1 Principle of Operation

Ever since the atomic structure of matter was envisioned, imaging individual atoms remained a cherished though elusive goal until the introduction of the STM. Gerd Binnig and Heinrich Rohrer developed the first working STM in 1981 while working at IBM Zurich Research Laboratories in Switzerland (Chen 2008). This humble instrument provided a breakthrough in human ability to investigate matter on the atomic scale: for the first time, the individual surface atoms of flat samples could be made visible in real space. This instrument later won Binnig and Rohrer the Nobel Prize in Physics in 1986.

STM works by the united application of several principles: quantum-mechanical tunneling (quantum-mechanical phenomenon where a particle tunnels through a barrier that it classically could not surmount because its total kinetic energy is lower than the potential energy of the barrier), piezo-electric effect (the ability of a material to produce electricity when subjected to mechanical stress) and feedback loop (in which information about the result of a transformation or an action is sent back to the input of the system in the form of input data allowing self-correction) (Bowker and Davies 2010). The spectacular spatial resolution of the STM along with its intriguing simplicity launched a broadscale research effort significantly impacting surface science. A large number of metals and semiconductors have been investigated on the atomic level and marvelous, beautiful images of the world of atoms were created within the first few years after the inception of the STM. Today, the STM is an invaluable asset in the toolbox of a surface analysis scientist.

Thinking in classical terms, a particle is not allowed to move over a barrier unless it has sufficient energy to cross the barrier. If two metallic objects are connected to a battery, one to each end and the circuit is completed by touching the two objects together, a current starts flowing between the objects. If the objects are pulled apart and separated, the current ceases to flow because of inability of electrons to cross the air gap between the objects. It is said that the air gap presents a "potential energy barrier" to the electrons that cannot be compensated by the battery voltage. But when distances are measured at atomic scales, classical rules are not applicable. In quantum mechanics, electrons have wavelike properties (Figure 1.15). A tunneling current occurs
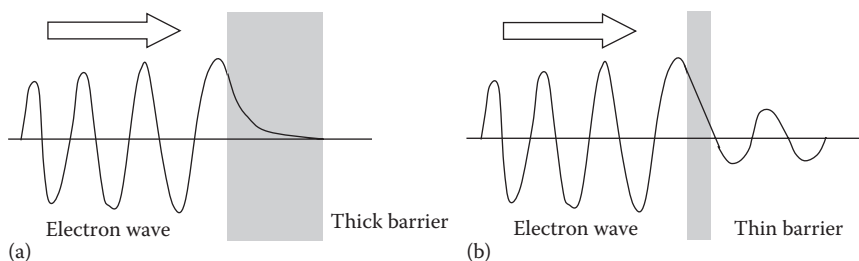
FIGURE 1.15
Interaction of electron wave with (a) thick potential barrier and (b) thin barrier. In (a), the wave does not end abruptly but tapers off exponentially. However, it does not cross the barrier. Something is said to increase or decrease exponentially if its rate of change is expressible by an exponential function. A graph of such a rate would appear as a curve, not as a straight line. In (b), the wave manages to penetrate the barrier and some electrons appear on the opposite side of the barrier.

when electrons move across a barrier that is classically impenetrable for them. This becomes obvious from the realization that the electron waves do not terminate abruptly at a wall or barrier, but taper off quickly on encountering a barrier. For a very thin barrier, the probability function extends into the next region through the barrier. Because of the small but nonzero probability of the presence of an electron on the opposite side of the barrier, if a large number of electrons is available on one side of the barrier, some electrons will successfully move through the barrier and make their appearance on the other side of the same. When electrons move through the barrier in this fashion, the phenomenon is called tunneling through the barrier, a case that is forbidden in classical physics.

The sample is scanned by a very fine metallic tip mechanically connected to the scanner, an *XYZ* positioning device realized by means of piezoelectric materials (Figure 1.16). The scanner allows 3D positioning in the *X, Y,* and *Z* directions with subatomic precision. The tunneling tip is typically a wire that has been sharpened by chemical etching or mechanical grinding. W, PtIr, or pure Ir is often chosen as the tip material. The sample is positively or negatively biased so that when the distance between tip and sample is in the range of several angstroms, a small current, the tunneling current, flows with the tip very close but not in contact with the sample. This feeble tunneling current is amplified and measured. This current is used as the feedback signal in a Z-feedback loop. With the help of the tunneling current I_t, the feedback electronics keeps the distance between tip and sample constant. In the topographic mode, images are created by scanning the tip in the *XY* plane and recording the Z position required to keep I_t constant. If the tunneling current exceeds its preset value, the distance between tip and sample is increased. When it falls below this value, the feedback decreases the distance. The tip is scanned line by line above the sample surface following the topography of the sample. In the *constant-height mode*, the probe scans rapidly so that the
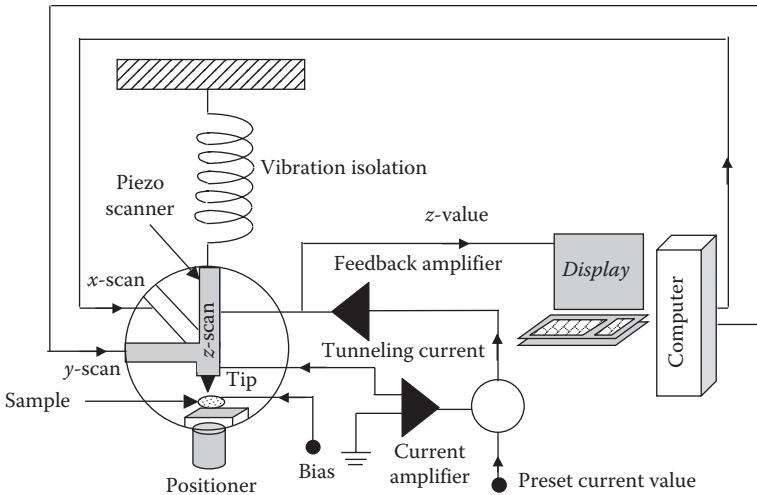
FIGURE 1.16
Scanning tunneling microscope. (After Chen, C. J., *Introduction to Scanning Tunneling Microscopy*, Oxford University Press, New York, 2008.)

feedback cannot follow the atomic corrugations. The atoms are then apparent as modulations of I_t, which are recorded as a function of X and Y. The scanning is usually performed in a raster fashion with a fast scanning direction (sawtooth or sinusoidal signal; named sawtooth based on its resemblance to the teeth on the blade of a saw) and a slow scanning direction (sawtooth signal). A computer controls the scanning of the surface in the XY plane while recording the Z position of the tip (topographic mode) or I_t (constant-height mode). Thus, a 3D image Z (X, Y, $I_t \sim$ const) or I_t (X, Y, $Z \sim$ const) is created.

1.19.2 Transmission Coefficient

Schrödinger's equation of quantum mechanics is applied to predict the increase of tunneling current between the tip and the sample with decrease in separation between two metals. Here, the energy of all the electrons in the metal is lower than the height of the potential wall. Let the difference between the most energetic electron and the vacuum energy be denoted by the symbol ϕ, which is called the work function of the material. The movement and shape of the electron wave is governed by Schrödinger's equation. Although, the complete description of the tunneling process requires a solution of the 3D form of Schrödinger's equation, a 1D analysis is sufficient to describe the main features of the phenomenon. The 1D Schrödinger equation is given by

$$\left(-\frac{\hbar^2}{2m}\right)\nabla^2\psi(x,t) + U(x)\psi(x,t) = i\hbar\left[\frac{\partial\psi(x,t)}{\partial t}\right] \tag{1.7}$$

where $U(x)$ is the potential energy of the electron as a function of its position. The plane wave (a constant-frequency wave whose wavefronts, that is, surfaces of constant phase, are infinite parallel planes of constant amplitude normal to the phase velocity vector) for an electron wave function of wave number

$$k = \frac{2\pi}{\lambda} \tag{1.8}$$

and angular frequency ω (a scalar measure of rotation rate in radians per second where radian is a unit of angle similar to degrees) is represented by the equation

$$\psi(x,t) = A\exp\{-i(kx + \omega t)\} + B\exp\{i(kx - \omega t)\} \tag{1.9}$$

Electrons of energy $E(x, t) = E$ bumping into a uniform potential barrier of height $U(x, t) = U(x)$ continuously flow from one metal to the other. A steady-state time-independent situation is postulated. Under this assumption, it is necessary to solve only the 1D steady-state Schrödinger equation, given by

$$\left(\frac{-\hbar^2}{2m}\right)\left\{\frac{\partial^2\psi(x)}{\partial x^2}\right\} + U(x)\psi(x) = E\psi(x) \tag{1.10}$$

where
  E is the kinetic energy of the electron
  $U(x)$ is the potential energy of the electron as a function of position, is smaller than the electron energy in the metals and larger than the electron energy in the barrier

For sake of simplicity, it is assumed that

$$U(x) = U_0 \tag{1.11}$$

a constant in the barrier. The general solutions to the above equation in the Metal 1, Metal 2, and barrier are given, respectively, by the equations (University of Wisconsin 2007)

$$\text{Metal 1}\quad \psi(x) = A\exp\{-(ikx)\} + B\exp\{(ikx)\}, \quad k = \frac{\sqrt{2m(E - U_0)}}{\hbar} \tag{1.12}$$

$$\text{Metal 2}\quad \psi(x) = E\exp\{-(ikx)\} + F\exp\{(ikx)\} \tag{1.13}$$

$$\text{Barrier}\quad \psi(x) = C\exp\{-(\mu x)\} + D\exp\{(\mu x)\}, \quad \mu = \frac{\sqrt{2m(U_0 - E)}}{\hbar} \tag{1.14}$$

Equations 1.12 and 1.13 show that the phase of the electron wave function undergoes uniform variation in the metals. The wavelength is $\lambda = 2\pi/k$. Higher-energy electrons have a larger frequency and smaller wavelength. When a high-energy electron wave stumbles upon the boundary of the metal, it "leaks out" a very small amount. The "intensity" of the electron wave declines as a function of distance from the boundary. Mathematically speaking, the argument of the exponential function becomes real and the electron wave function decays. (For imaginary arguments, the wave function will have oscillatory behavior.)

   As the objective is to acquire a quantitative insight into the electron tunneling phenomenon, it is necessary to derive an expression for the transmission coefficient (the amplitude or the intensity of a transmitted wave relative to an incident wave), which indicates the transmitted flux from the sample to the tip through the barrier of width L. The barrier is considered to be wide but finite, such that the electron wave function decays in the barrier significantly. Furthermore, the energy and mass conservation requirements impose the condition that electron wave function and its first derivative must be continuous, that is, join smoothly at the sample-barrier and tip-barrier boundaries. Let us set up a coordinate system in which the surface of the sample (Metal 1) is at $x = 0$ and the tip (Metal 2) is at $x = L$. Then applying the boundary conditions (the set of conditions specified for the behavior of the solution to a set of differential equations at the boundary of its domain) for continuity, we get

$$A + B \approx C \tag{1.15}$$

$$ik(A - B) \approx -\mu C \tag{1.16}$$

(at the sample surface, $x = 0$) where D, the amplitude of the reflected wave function at the tip–barrier boundary, is neglected, since $D \ll A, B, C$. However, D is not small to be *negligible* at the tip–barrier boundary. At the tip–barrier boundary, $x = L$, continuity requires the following:

$$C \exp\{-(\mu L)\} + D \exp\{(\mu L)\} = F \exp\{(ikL)\} \tag{1.17}$$

$$-\mu C \exp\{-(\mu L)\} + \mu D \exp\{(\mu L)\} = ikF \exp(ikL) \tag{1.18}$$

Solving for B/A at $x = 0$, by obtaining the solution for C and substituting for it, we have

$$\frac{B}{A} = -\frac{1 + ik\delta}{1 - ik\delta} \tag{1.19}$$

where

   δ is $1/\mu$

   A is the amplitude of the electron wave function in the sample surface incident on the barrier

   B represents the amplitude of the reflected wave function

The reflection coefficient (*R*: the ratio of the amplitude of a wave reflected from a surface to the amplitude of the incident wave) for the wave function is then defined as

$$R \oplus \left| \frac{B}{A} \right|^2 \tag{1.20}$$

Physically, it represents the relative intensities of the incident and reflected wave functions.

   For an electron incident on the barrier, there are two possibilities: It is either reflected or transmitted through the barrier. In terms of probability or frequency of occurrence, $R + T = 1$, where *R* and *T* are the reflection and transmission coefficients. Thus,

$$R = \left| \frac{B}{A} \right|^2 = \left| -\frac{1 + ik\delta}{1 - ik\delta} \right|^2 \approx 1 \tag{1.21}$$

and therefore

$$T = 1 - R \approx 0 \tag{1.22}$$

which indicates that, for an infinitely wide barrier, the number of electrons found in the barrier region will be zero. Nevertheless, division of the first of the sample vacuum-barrier boundary conditions by *A* results in the equation

$$1 + \frac{B}{A} \approx \frac{C}{A} \tag{1.23}$$

The probability of finding an electron in the barrier region at $x = 0$, due to quantum-mechanical tunneling, is given by the equation

$$\left| \frac{C}{A} \right|^2 \approx \frac{4(k\delta)^2}{1 + (k\delta)^2} \tag{1.24}$$

For determination of the effective tunneling transmission coefficient, F/A^2, that is, the relative probability for frequency of occurrence of an electron

tunneling out of the sample surface, across the sample–tip–barrier region, and into the tip, the tip–barrier boundary equations (at $x=L$) and Equation 1.18 are combined to get

$$\left|\frac{F}{A}\right| \approx -\frac{4ik\delta}{(1-ik\delta)^2}\exp\left[-L(\mu+ik)\right] \tag{1.25}$$

which produces the desired quantitative result

$$T(E) = \left|\frac{F}{A}\right|^2 \approx \left[\frac{4k\delta}{\{1+(k\delta)^2\}}\right]^2 \exp\left(\frac{-2L}{\delta}\right) \propto \exp\left(-2L\sqrt{\frac{2m\Phi}{\hbar^2}}\right) = \exp\left(-\frac{L}{L_0}\right) \tag{1.26}$$

where

$$k^2 = \frac{2mE}{\hbar^2}, \quad (k\delta)^2 = E(U_0 - E) = E\Phi \quad \text{and} \quad L_0 = \frac{\hbar}{\sqrt{8m\Phi}} \tag{1.27}$$

Substituting typical values: $\Phi = 5\times10^{-19}$ J, $m = 9.11\times10^{-31}$ kg, and $\hbar = 1.05\times10^{-34}$ J-s, results in a decay length $L_0 = 0.55$ Å so:

$$T(E) \approx \exp(-2L) \tag{1.28}$$

where L is expressed in angstroms. This formula shows that the probability of tunneling of electron across the barrier decreases by an order of magnitude for each angstrom change in separation. This demonstrates mathematically that tunneling current is an extremely sensitive measure of the distance between the tip and sample.

Example 1.3

Determine the tunneling transmission coefficients $T(E)$ for barrier widths (L) of 1, 2, 3, 5, 8, and 10 Å and plot T as a function of L.
From Equation 1.28, $T(E) = \exp(-2\times1) = \exp(-2) = 0.135$ for $L = 1$ Å. Similarly, for $L = 2$, 3, 5, and 10 Å, $T(E) = 0.0183$, 0.00248, 4.54×10^{-5}, and 2.06×10^{-9}, respectively. $T(E)$ versus L plot is shown in Figure 1.17.

In the STM, one of the metals is the sample being investigated and the other metal is the probe. The sample is usually flatter than the probe. Because the probe is formed of atoms, if it is sharpened into a tip, it will most likely have one atom at the end of the tip. The spacing between atoms is about 3 Å. Therefore, any tunneling through atoms that are one atom back from the closest atom is a fraction exp (–2)(3) of tunneling through the atom at
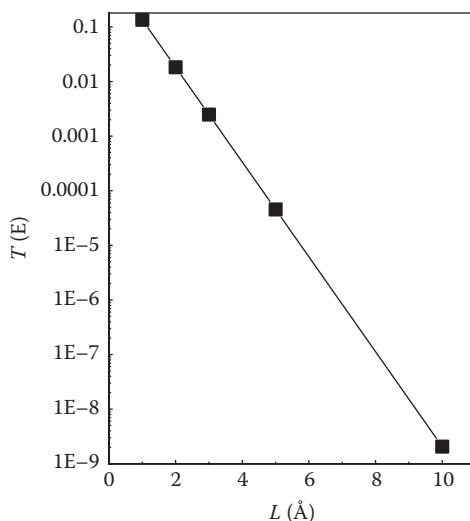
FIGURE 1.17
Effect of tip–sample spacing on transmission coefficient.

the tip. Virtually all of the tunneling electrons will pass through the single atom closest to the surface. Most of the tunneling current is carried by the atom that is closest to the sample (the "front atom"). If the sample is very flat, this front atom remains the atom that is closest to the sample during scanning in X and Y, and even relatively blunt tips yield atomic resolution easily. The sample surface is scanned with a single atom! This feature produces the atomic resolution capabilities of the microscope. STM, in fact, still provides the best resolution available. STM resolution can be as good as 0.1 nm. Z resolution is about 0.1 nm for a well-designed STM.

The spectacular spatial resolution and relative ease of obtaining atomic resolution by scanning tunneling microscopy is based on three properties of the tunneling current: (i) As a consequence of the strong distance dependence of the tunneling current, even with a relatively blunt tip, the chance is high that a single atom protrudes far enough out of the tip that it carries the main part of the tunneling current. (ii) Typical tunneling currents are in the nanoampere range—measuring currents of this magnitude can be done with a very good signal-to-noise ratio even with a simple experimental setup. (iii) Because the tunneling current is a monotonic function (a function that is either entirely nonincreasing or nondecreasing) of the tip–sample distance, it is easy to establish a feedback loop that controls the distance so that the current is constant.

Although the STM itself does not need vacuum to operate (it works in air as well as under liquids), ultrahigh vacuum is required to avoid contamination of the samples from the surrounding medium. Unfortunately, there are some limitations, the most important of which is that the surface of the

sample must be conducting or semiconducting. This limits the materials that can be studied by STM. This limitation led to the invention in 1986 of the first AFM. The STM can only image electrically conductive samples, which limits its application to the imaging of metals and semiconductors. But even conductors—except for a few special materials, like highly oriented pyrolytic graphite (HOPG: a unique form of graphite grown on to a substrate by decomposition of a hydrocarbon gas at very high temperature in a vacuum furnace, and permitting the graphite to crystallize; the result is an ultrapure product that is near theoretical density and extremely anisotropic)—cannot be studied in ambient conditions by STM but have to be investigated in an ultrahigh vacuum. In ambient conditions, the surface layer of solids constantly changes by adsorption and desorption of atoms and molecules. An ultrahigh vacuum is required for clean and well-defined surfaces. Because electrical conductivity of the sample is not required in atomic force microscopy, the AFM can image virtually any flat solid surface without the need for surface preparation. Consequently, thousands of AFMs are in use in universities and public and industrial research laboratories all over the world. Most of these instruments are operated in ambient conditions.

1.19.3 Tunneling Current

The tunneling current flowing between the tip and the sample depends on the voltage difference between these bodies. On biasing the sample negatively by a voltage $-V$ with respect to the tip, the Fermi level of the sample electrons with respect to the tip electrons is effectively raised and electrons tend to flow out from the filled states of the sample into the empty states of the tip. For states of energy E (with respect to the Fermi level of the sample), the elastic tunneling current from the sample to the tip is expressed as the product of several factors (Hoffman Lab 2010a):

$$I_{\text{sample}-\text{tip}} = -2E \times \left(\frac{2\pi}{\hbar}\right) \left|\left(\frac{F}{A}\right)\right|^2 \times \rho_s(E)f(E) \times \left[\rho_t(E+qV)\left\{1 - f(E+qV)\right\}\right] \quad (1.29)$$

where
  the factor $\rho_s(E)f(E)$ represents the filled sample states for tunneling from
    the sample
  the factor $[\rho_t(E+qV)\{1 - f(E+qV)\}]$ is the empty tip states for tunneling to
    the tip
  the factor of 2 accounts for electron spin
  $-q$ is the electron charge
  $|(F/A)|^2$ is a matrix element
  $f(E)$ is the Fermi distribution function

$$f(E) = \left\{ 1 + \exp\left(\frac{E}{k_\mathrm{B}T}\right)\right\}^{-1} \tag{1.30}$$

Though the predominant tunneling current flow for negative sample voltage $-V$ is directed from sample to tip, there exists a smaller reverse tunneling current of electrons from tip to sample. The governing equation for this current is written in an identical manner. By summation of these currents, and integration over all energies E, the total tunneling current from sample to tip becomes

$$I_\mathrm{total} = -\left(\frac{4\pi E}{\hbar}\right)\int_{-E_\mathrm{F}}^{\infty}\left|\left(\frac{F}{A}\right)\right|^2 \rho_\mathrm{s}(E)\rho_\mathrm{t}(E + qV)$$

$$\times \left[f(E)\{1 - f(E + qV)\} - \{1 - f(E)\}\{f(E + qV)\}\right]dE \tag{1.31}$$

This basic equation can be simplified by applying various realistic approaches to obtain meaningful solutions. First, suppose the measurements are performed at low temperatures. In this case, the Fermi function is sharply cut off at the Fermi surface with a cutoff width of $k_\mathrm{B}T$, which is only 0.36 meV for $T = 4.2$ K. In the approximation of an ideally abrupt cutoff, the integral is subdivided into three parts, named as I, II, and III. Therefore, for finding the tunneling current, the relevant range of E over which integration must be performed is reduced to region II only, where $-qV < E < 0$. (Likewise, if a positive bias voltage V had been applied to the sample, the range of integration is $0 < E < qV$.) This results in considerable saving of effort. Hence, approximately

$$I_\mathrm{total} \approx -\left(\frac{4\pi E}{\hbar}\right)\int_{-qV}^{0}\left|\left(\frac{F}{A}\right)\right|^2 \rho_\mathrm{s}(E)\rho_\mathrm{t}(E + qV)dE \tag{1.32}$$

Second, let us choose a tip material having a flat density of states (DOS; the number of states per interval of energy at each energy level that are available to be occupied) within the energy range of the Fermi surface (an abstract constant–energy interface that defines the allowable energies of electrons in a solid; it is useful for characterizing and predicting the thermal, electrical, magnetic, and optical properties of crystalline metals and semiconductors) to be studied. As an example, if the sample DOS is to be studied within 200 meV of the Fermi surface, then the measured tunneling current is a convolution (a mathematical operation on two functions, producing a third function that is typically viewed as a modified version of one of the original functions; an

integral that expresses the blending or amount of overlap of one function as it is shifted over another function) of the DOS of the tip and sample in this energy range. Then a tip material is selected such that it has a flat DOS in this range, so that $\rho_t(E + qV)$ is treated as a constant and taken outside the integral.

$$I_{\text{total}} \approx \left(\frac{4\pi E}{\hbar} \right) \rho_t(0) \int_{-qV}^{0} \left| \left(\frac{F}{A} \right) \right|^2 \rho_s(E) dE \tag{1.33}$$

From the basic theory for vacuum tunneling, under realistic assumptions, the following is evident: (i) The tip and the sample each have their own independent DOS. (ii) Each of their wave functions falls exponentially to zero in the tunneling barrier. (iii) The overlapping is small enough (i.e., tip–sample separation is sufficiently large) that each side is feebly influenced by the tail of the wave function from the other side. Under this condition, the matrix element for tunneling is independent of the energy difference between the two sides of the barrier. To a reasonable approximation, the matrix element is taken outside the integral and treated as a constant.

$$I_{\text{total}} \approx \left(\frac{4\pi E}{\hbar} \right) \left| \left(\frac{F}{A} \right) \right|^2 \rho_t(0) \int_{-qV}^{0} \rho_s(E) dE \tag{1.34}$$

But the matrix element arises from the assumption that both tip and sample wave functions fall off exponentially into the vacuum gap. Basically, a square vacuum barrier is assumed, and a WKB (Wentzel–Kramers–Brillouin) approximation (a method for finding approximate solutions to linear partial differential equations with spatially varying coefficients) is done. In reality, there will be some tilting to the top of the barrier, but the tilt will be the applied voltage (around 100 meV) while the height of the barrier equals the energy required to remove an electron from a metal, that is, the work function, has a value of several electron volts. Therefore, the tilt of the barrier will be much smaller than the height of the barrier, and can be ignored. According to WKB, the tunneling probability through a barrier is

$$\left| \left(\frac{F}{A} \right) \right|^2 = \exp(-2\gamma) \tag{1.35}$$

where

$$\gamma = \int_{0}^{s} \sqrt{\frac{2m\varphi}{\hbar^2}} \, dx = \frac{s}{\hbar} \sqrt{2m\varphi} \tag{1.36}$$

m is the mass of the electron

s is the width of the barrier (tip–sample separation)

φ is the height of the barrier, which is actually a combined effect of the work functions of the tip and sample

The work function is measured by recording the tunneling current as a function of tip–sample separation.

$$I \propto \exp\left(-\frac{2s}{\hbar\sqrt{2m\varphi}}\right) \tag{1.37}$$

Therefore, φ is obtained from the slope of the natural logarithm of tunneling current versus barrier width, that is, I versus s graph. Typically, φ is ~ 3–4 eV. The higher the value of φ, the more the tunneling current I varies for a given change in s; therefore, a higher φ provides a superior resolution tip.

However, due to the exponential falloff, measurement of the absolute value of s is not possible. This is troublesome, because there is no way to ensure that measurements are being made at constant tip–sample separation. So, if variations are noticed from one point on the sample surface to another, it is not sure whether the variation is due to intrinsic inhomogeneities in the sample at the specific energy of measurement, or due to a varying tip–sample separation, that is, it is difficult to distinguish between the two causes.

In summary, the tunneling current is fairly well approximated by the equation

$$I_{total} \approx \left(\frac{4\pi q}{\hbar}\right)\exp\left(-s\sqrt{\frac{8m\phi}{\hbar^2}}\right)\rho_t(0)\int_{-qV}^{0}\rho_s(E)dE \tag{1.38}$$

1.19.4 Measurements with STM

1.19.4.1 Topography

The common mode of STM measurement employed by research groups worldwide is "topography" (Hoffman Lab 2010b). In this mode, the tip is rastered across the surface at a fixed sample bias voltage V_{set}. A feedback loop is employed for controlling the voltage across the z piezoelement for maintaining a constant value of the tunneling current I_{set}. By recording the voltage applied to the z piezo, the height of the surface is effectively mapped.

The meaning of the "height of the surface" is unclear. One obvious suggestion is that some contour of constant charge density is implied. However, as can be seen from Equation 1.38 for I_{total}, the tunneling current is not dependent on the total charge density, but depends only on the charge density within qV below the Fermi surface, where $-V$ is the applied bias.

An arbitrarily large applied voltage will enable more charge density to be captured, but there are two problems: (i) Some of the samples are fragile compounds with weak bonds, so that if a large voltage is applied locally, pieces of the surface will literally rip off. (ii) If V is too high (~ the work function φ), the tunneling approximation is invalid.

So the "height of the surface" is somewhat arbitrarily defined as the tip–sample separation for which tunneling current is fixed at a particular constant value I_{set} for a particular applied bias voltage V_{set}. Practically, a fixed current is chosen at −100 pA, for a bias voltage of −100 mV. This is arbitrary but gains support from the fact that in accordance with expectations, atoms and other structural features are not seen, even over a wider range of choices of I_{set} and V_{set}. The most widely varying DOS features of the superconducting samples studied so far appear to be within 75 meV of the Fermi level. Superconductivity is occurrence of zero electrical resistance in certain materials below a characteristic temperature.

1.19.4.2 Density of States

From the tunneling equation mentioned earlier, it is easy to see that if the tip–sample separation is held constant, at a given (x, y) location, and a negative bias voltage $−V$ is applied on the sample, we have

$$I = I_0 \int_{-qV}^{0} \rho_s(E)dE \qquad (1.39)$$

In other words, we can measure the integral of the DOS, down to any energy $−qV$, by varying $−V$. For a negative bias voltage on the sample, electrons are tunneling from sample to tip, and the integrated density of full states below the Fermi level in the sample is being measured. For a positive bias voltage on the sample, electrons are tunneling from tip to sample, and the integrated density of empty states above the Fermi level in the sample is measured. Thus, the integrated density of states (IDOS) is obtained. But it would be useful to find the DOS. After plotting an IDOS versus V curve, a numerical derivative (a technique of numerical analysis to produce an estimate of the derivative of a mathematical function using values from the function and other knowledge about the same) of the data is taken to get the DOS. But it will be easier to measure the derivative directly. So a lock-in amplifier (a type of amplifier that can recover signals in the presence of an overwhelming noise background or can provide high-resolution measurements of relatively clean signals over several orders of magnitude) is employed to modulate the bias voltage by dV (typically a few millivolts) around a DC voltage V of interest. Due to the voltage modulation dV, a current modulation dI can be measured. This dI/dV is termed the conductance $g(V)$, so that we can write

$$g(V) \equiv \frac{dI}{dV} \propto \text{DOS (eV)} \tag{1.40}$$

Therefore, by using a lock-in amplifier and varying V, an entire DOS curve can be mapped.

The energy resolution is limited by the amplitude of the wiggle until the modulation becomes less than approximately $K_B T = 0.36 \text{meV}$ at $T = 4.2 \text{K}$. So ideally, the voltage modulation can be made smaller than 0.36 mV. But practically, adequate signal-to-noise ratio (the dimensionless ratio of the signal power to the noise power corrupting the signal; it provides a comparison of the amount of a particular signal with the amount of background noise) cannot be obtained at this low amplitude without prohibitively long averaging times. Most of the data is measured with a 2 mV RMS modulation, therefore blurring the energy resolution by approximately 5.6 meV.

1.19.4.3 Linecut

In the previous section, a single DOS curve at a single location was discussed. The possibility of (x, y) control over the location of the tip using the piezo tube scanner enables the measurement of DOS curves at any desired location. Some samples, like good metals (without impurities), will have a completely homogenous DOS everywhere. But other more interesting samples are inhomogeneous. A full DOS curve can be measured at every point along a straight line, spaced a few angstroms apart, and a "linecut" can be seen.

1.19.4.4 DOS Map

Basically, a 3D data set is being discussed: two spatial dimensions x and y (by varying the position of the tip) and one energy dimension (by varying V). This 3D data set can be viewed as a series of DOS-versus-energy curves at every location (x, y), or as a series of 2D DOS-maps at each energy qV. Mapping the DOS at a specific energy is a good visual representation to see the inhomogeneities in the DOS.

1.20 The Force Nanosensor: AFM

1.20.1 Operating Principle

AFM is the most versatile member of the family of scanning probe microscopes (SPMs) (Morris et al. 2004). The AFM is closely related to the STM, and it shares its key components, except for the probe tip. X and Y topographic

resolution for most AFMs is typically 2–10 nm. The AFM works in the same way as our fingers touch and probe the environment when we cannot see things in darkness. By using fingers to "visualize" an object, our brain is able to deduce its topography while touching it. *The AFM generates images by feeling the specimens as opposed to seeing by optical microscopes.* Its closest predecessor is the stylus profiler, an instrument used to measure the profile of a surface, in order to quantify its roughness, step height, or thin film thickness. AFM technology uses sharper probes and lower forces than stylus profilers to provide higher-resolution information without sample damage. The essential part of AFM is a silicon (Si) or silicon nitride (Si_3N_4) cantilever with a sharp tip at its end having a tip radius of the order of nanometers. AFM works by bringing the cantilever tip in contact with the surface to be imaged. The AFM is similar to an STM, except that the tunneling tip is replaced by a force sensor. The potential energy between the tip and sample V_{ts} causes a Z component of the tip–sample force $F_{ts} = -\partial V_{ts}/\partial z$ and a tip–sample spring constant $k_{ts} = -\partial F_{ts}/\partial z$. Depending on the mode of operation, the AFM uses F_{ts} or some entity derived from F_{ts} as the imaging signal. Spring constant is the restoring force of a spring per unit of length. It is given by the change in the force exerted by the spring, divided by the change in deflection of the spring, that is, the gradient of the force versus deflection curve.

Unlike the tunneling current, which has a very short range, F_{ts} has long- and short-range contributions. We can classify the contributions by their range and strength. In vacuum, there are short-range chemical forces (fractions of nm) and van der Waals, electrostatic, and magnetic forces with a long range (up to 100 nm). In ambient conditions, meniscus forces formed by adhesion layers on tip and sample (water or hydrocarbons) can also be present. Meniscus is the free surface of a liquid, which assumes a flat, convex or concave shape, depending on the solid and liquid surface. The forces acting on a liquid molecule at the free surface are (i) the weight of the molecule, acting vertically downward; (ii) the force of adhesion; and (iii) the force of cohesion. Cohesion is the term for molecules of a substance sticking together. Liquid molecules are not only attracted to each other, but to any molecule with positive or negative charges. When a molecule attracts to a different substance, it is termed "adhesion."

1.20.2 Lennard–Jones Potential and the van der Waals Forces

Before describing the operation of AFM, it is essential to understand the fundamental mechanisms responsible for interactions between particles. A simple mathematical description of the interaction between two particles is given by the Lennard–Jones potential, (also referred to as the L–J potential, 6–12 potential or, less commonly, 12–6 potential). The Lennard–Jones potential function is a reasonably accurate model of interactions between noble gas atoms. The Lennard–Jones potential (Figure 1.18) is given by the expression
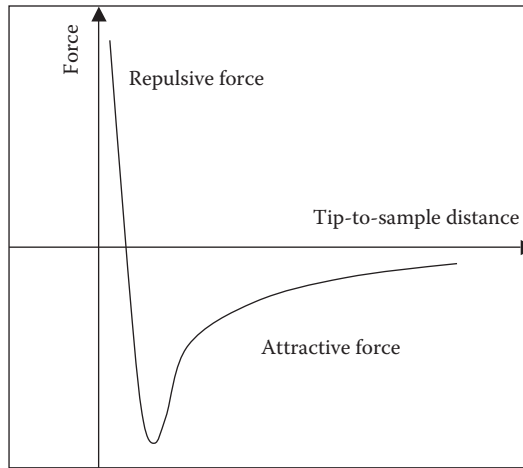
FIGURE 1.18
Graph of Lennard–Jones potential.

$$\Phi_{LJ}(r) = 4\varepsilon \left\{ \left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right\} \tag{1.41}$$

for interaction between a pair of atoms. Choice of parameters ε and σ is done to fit the physical properties of the material. The binding energy ε is the depth of the potential well. σ is the (finite) distance at which the interparticle potential is zero, and r is the distance between the particles.

This potential has an attractive tail at large r. It reaches a minimum around 1.122σ, and becomes strongly repulsive at shorter distances, passing through 0 at $r = \sigma$ and increasing steeply as r is further decreased. The term $\sim 1/r^{12}$, dominating at short distances, models the repulsion between atoms when they approach very close to each other. Its physical origin is related to the Pauli's exclusion principle, namely, no two electrons in an atom can be at the same time in the same quantum state or configuration, that is, they cannot have the same set of four quantum numbers; this prevents matter from collapsing to an extremely dense state. As soon as the electronic clouds surrounding the atoms start overlapping, the energy of the system increases abruptly. In the Lennard–Jones potential expression, the exponent 12 was selected solely on a practical basis. In fact, an exponential behavior is more appropriate on physical grounds.

Thus, the interaction between neutral atoms and molecules can be decomposed into two different forces: an attractive force at long distances (the van der Waals force) and a repulsive force at short distances (due to overlap between electron wave functions). Two interacting neutral atoms are subject to two opposing forces: first, they are weakly attracted by van der Waals

forces; second, they are repelled by Pauli repulsion. It is known that the van der Waals forces decay proportional to the 6th power of the separation and that the effects of the Pauli repulsion decay exponentially. The exponent 12 was chosen exclusively because of ease of computation. This form of the potential has little theoretical justification, but sometimes matches reality to an acceptable standard. Its use greatly simplifies numerical calculations based on the interatomic potential.

Elaborating on the attractive part, two factors contribute to this part:

1. *Dipole–dipole interaction*, that is, electrostatic attraction between two molecules with permanent dipole moments. An electric dipole is a separation of positive and negative charges. The dipole moment is defined as the product of the total amount of positive or negative charge and the distance between their centroids. If both the molecules have permanent dipoles, they will tend to align directions so as to produce a relatively strong attraction. Thus, the main contribution to the force between the two molecules will not depend on momentary induced fluctuations. This is called a permanent dipole–permanent dipole force. It is an electrostatic interaction and decreases with the 4th power of distance

$$F_{\text{permanent dipole}} = -\text{constant} \times R^{-4} \tag{1.42}$$

   where the constant of proportionality just depends on the identity of the molecules involved.

2. *Dipole-induced dipole interactions*, in which the dipole of one molecule polarizes a neighboring molecule, that is, the presence of one dipole (whether permanent or due to random fluctuation of charge density) exerts a force on the electrons of the other molecule causing a brief shift of orbital electrons to one side of one atom or molecule, creating a similar shift in adjacent atoms or molecules, thus giving it in turn a temporary dipole moment. Thus, weak attractive forces exist between atoms or nonpolar molecules. Essentially, the attraction between the molecules seen as electric dipoles is attraction between electron-rich regions of one molecule and electron-poor regions of another. These forces, called *London forces* or *dispersion forces*, hold together molecules with no permanent dipole moment.

The dispersion force is the weakest intermolecular force. It is a temporary attractive force that results when the electrons in two adjacent atoms occupy positions that make the atoms form temporary dipoles. It is sometimes called dipole-induced dipole attraction. This attractive force causes nonpolar substances to condense to liquids and to freeze into solids when the temperature is lowered sufficiently.

Clarifying further, the dispersion effect is the interaction between the instantaneous dipoles formed in the atoms by their orbiting electrons. The very rapidly changing dipole of one atom produces an oscillating electric field that acts upon the polarizability of a neighboring atom. The microscopic electric polarizability of an atom refers to its ability to respond to an external electric field by shifting its charges so as to create a dipole oriented favorably in the external field (negative end of the dipole closer to the positive pole of the external field). For an isolated atom this may occur by biasing the time-average distribution of electrons among ground and excited state to favor polar states. If excitation is easy, polarizability is high. For molecules in the gas or liquid state, skeletal vibrational states and rotational states may also contribute to the polarizability. Molecules that bend easily to produce a dipolar configuration have high polarizability. In the solid state, the rotational states are usually suppressed. If the atoms exist in a vacuum, the induced dipole of the neighboring atom moves in phase with the original dipole, producing an attractive atomic interaction.

Dispersion forces can also be thought of as very weak intermolecular covalent bonds, resulting from momentary dipoles forming due to asymmetry in electron cloud density. This electron cloud polarization causes a perturbation of electron clouds of nearby molecules, and the interaction of these temporary dipoles results in a net attractive interaction. Were it not for these intermolecular forces, molecules with no permanent dipole moment would exist solely in the gaseous phase.

These weak interactions dominate the bonding character of closed-shell systems, that is, rare gases such as Ar or Kr. Molecules such as dinitrogen (the diatomic molecule of nitrogen: N_2), methane (CH_4), and fullerene (a closed graphitic form of carbon with exactly 12 pentagons) are held in liquid and solid states by *dispersion forces*.

These forces (which also in fact occur between permanent dipoles) are weaker than permanent dipole forces and decrease with the 7th power of distance

$$F_{\text{induced dipole}} = -\text{constant} \times R^{-7} \tag{1.43}$$

The (different) constant of proportionality again depends on the molecules involved.

Curiously, the inverse 7th power van der Waals forces turn out to be very closely related to the Casimir effect, which tells us that a force acts between two close, parallel, uncharged conducting plates in a vacuum, due to quantum fluctuations in the electromagnetic field.

Thus, the forces acting between tip of the AFM cantilever and the sample surface are attractive or repulsive forces between molecular entities (or between groups within the same molecular entity) other than those due to bond formation or to the electrostatic interaction of ions or of ionic groups with one another or with neutral molecules. The attraction is much weaker

than a chemical bond. At short distances, the forces are repulsive in nature, which prevents the collapsing of molecules. As entities move closer to one another these repulsions dominate.

Example 1.4

An oxygen atom has an atomic radius of 60 pm. Apply the equation for Lennard–Jones potential to show that the interatomic force changes from repulsive to attractive as the interatomic distance increases.

Considering that the distance between two oxygen atoms is minimum when distance between the centers of two oxygen atoms = atomic diameter = 120 pm, for the closest approach of two oxygen atoms $\sigma = 120$ pm. At $r < 120$ pm, σ/r is >1. Then, the first term in Lennard–Jones potential is larger than the second term because the 12th power of a number >1, has a greater magnitude than the 6th power of the same number. Consequently, the potential has a positive value corresponding to repulsive force. But for $r > 120$ pm, σ/r is <1, that is, a fraction. Twelfth power of a fractional number is smaller than its 6th power. First term of L–J potential is smaller than the second term giving negative value of potential; hence, the force is attractive. Potential energy is relative to some datum or reference level where the potential energy is zero here at $r = \sigma$. If one extracts potential energy from the charge by letting it move with the attractive force, then the potential energy becomes smaller than the datum value, and hence negative. If one adds work to the charge to move it against the repulsive forces (from fields), its potential energy increases with respect to that datum and the potential energy is positive.

Assuming additivity and disregarding the discrete nature of matter by replacing the sum over individual atoms by an integration over a volume with a fixed number density of atoms, the van der Waals interaction between macroscopic bodies can be calculated by the Hamaker approach (Hamaker 1937). This approach does not account for retardation effects due to the finite speed of light and is therefore only appropriate for distances up to several hundred angstroms. For a spherical tip with radius R next to a flat surface (z is the distance between the plane connecting the centers of the surface atoms and the center of the closest tip atom), the van der Waals potential is given by (Israelachvili 1991)

$$V_{vdW} = -\frac{A_H R}{6z} \qquad (1.44)$$

The Hamaker constant A_H depends on the type of materials (atomic polarizability and density) of the tip and sample. For most solids and interactions across a vacuum, A_H is of the order of 1 eV. Atomic polarizability is the electric dipole moment induced in a system, such as an atom or a molecule, by an electric field of unit strength.

1.20.3 Other Forces and Potentials

The force between tip and sample is composed of many contributions: electrostatic, magnetic, van der Waals, and chemical forces in vacuum. In ambient conditions, there are also meniscus forces. While electrostatic, magnetic, and meniscus forces can be eliminated by equalizing the electrostatic potential between tip and sample, using nonmagnetic tips, and operating in vacuum, the van der Waals forces cannot be switched off. For imaging by AFM with atomic resolution, it is desirable to filter out the long-range force contributions and measure only the force components that vary at the atomic scale.

A prototype of the chemical bond is treated in many textbooks on quantum mechanics: the H_2^+ ion is a model for the covalent bond. This quantum-mechanical problem can be solved analytically and gives interesting insights into the character of chemical bonds. The Morse potential (Israelachvili 1991), a convenient model for the potential energy of a diatomic molecule, describes a chemical bond with bonding energy E_{bond}, equilibrium distance σ, and a decay length κ. While the Morse potential can be used for a qualitative description of chemical forces, it lacks an important property of chemical bonds: anisotropy. Chemical bonds, especially covalent bonds, show an inherent angular dependence of the bonding strength. Using the Stillinger–Weber potential (one of the first attempts to model a semiconductor with a classical model; it is based on a two-body term and a three-body term), one can explain subatomic features in Si images. With increasing computer power, it becomes more and more feasible to perform ab initio calculations for tip–sample forces.

1.20.4 Force Sensor (Cantilever) and Force Measurement

The central element of a force microscope and its major instrumental difference from a STM is the spring that senses the force between tip and sample. For sensing normal tip–sample forces, the force sensor should be rigid in two axes and relatively soft in the third axis. This property is fulfilled with a cantilever beam, and, therefore, the cantilever geometry is typically used for force detectors. For a rectangular cantilever with dimensions w (width), t (thickness), and L (length), the spring constant k is given by (Chen 2008)

$$k = \frac{Ywt^3}{4L^3} \tag{1.45}$$

where Y is Young's modulus. The fundamental eigenfrequency (vibration frequency of a quantum system) f_0 is given by (Chen 2008)

$$f_0 = 0.162\left(\frac{t}{L^2}\right)\sqrt{\frac{Y}{\rho}} \tag{1.46}$$

where ρ is the mass density of the cantilever material. Force is calculated from the amount of bending of the cantilever measured by a laser spot reflected on to a split photo detector. It is calculated by Hooke's law by measuring the deflection of the cantilever and from knowledge of the spring constant of the material of the cantilever. The properties of interest are the stiffness (k: the resistance of an elastic body to deformation by an applied force, measured by the ratio of a steady force acting on the body to the resulting displacement), the eigenfrequency f_0, the quality factor (Q: parameter that describes how under-damped an oscillator is; a higher Q value indicates a lower rate of energy loss relative to the stored energy of the oscillator), the variation of the eigenfrequency with temperature $\partial f_0/\partial T$, and the chemical and structural composition of the tip.

The first AFMs were mostly operated in the static contact mode (see following text), and for this mode the stiffness of the cantilever should be less than the interatomic spring constants of atoms in a solid, which amounts to $k \leq 10\,\mathrm{N\ m^{-1}}$. This constraint on k was assumed to hold for dynamic atomic force microscopy as well. In the dynamic mode, the cantilever is externally oscillated at, or close to, its fundamental resonance frequency or a harmonic. However, it turned out later that in dynamic atomic force microscopy, k values exceeding hundreds of N m^{-1} help to reduce noise and increase stability. The Q factor depends on the damping mechanisms present in the cantilever. For micromachined cantilevers operated in air, Q is mainly limited by viscous drag (force arising from viscous effect of a fluid, that is, due to resistance to flow) and typically amounts to a few hundred, while in vacuum, internal and surface effects in the cantilever material are responsible for damping and Q reaches hundreds of thousands. The most common cantilevers in use today are built from all-silicon with integrated tips pointing in a [001] crystal direction. Self-sensing cantilevers with integrated tips have a built-in deflection-measuring scheme utilizing the piezoresistive effect in silicon. In dynamic atomic force microscopy, some requirements for the force sensor are similar to the desired properties of the time-keeping element in a watch: utmost frequency stability over time and temperature changes and little energy consumption.

Quartz tuning forks (small two-pronged devices) have many attractive properties, but their geometry gives them marked disadvantages for use as force sensors. The great benefit of the fork geometry is the high Q factor, which is a consequence of the presence of an oscillation mode in which both prongs oscillate opposite to each other. The dynamic forces necessary to keep the two prongs oscillating cancel in this case exactly. However, this only works if the eigenfrequency of both prongs matches precisely. The mass of the tip mounted on one prong and the interaction of this tip with a sample breaks the symmetry of tuning fork geometry. This problem can be avoided by fixing one of the two beams and turning the fork symmetry into a cantilever symmetry, where the cantilever is attached to a high-mass substrate with a low-loss material.

For atomic-resolution AFM, the front atom of the tip should ideally be the only atom that interacts strongly with the sample. In order to reduce the forces caused by the shaft of the tip, the tip radius should be as small as possible. Cantilevers made of silicon with integrated tips are typically oriented so that the tip points in the [001] crystal direction. Due to the anisotropic etching rates of Si and SiO_2, these tips can be etched so that they develop a very sharp apex. Not only the sharpness of a tip is important for atomic force microscopy, but also the coordination of the front atom. Tip and sample can be viewed as two giant molecules. In chemical reactions between two atoms or molecules, the chemical identity and the spatial arrangement of both partners play a crucial role. For AFM with true atomic resolution, the chemical identity and bonding configuration of the front atom is therefore critical. In [001]-oriented silicon tips, the front atom exposes two dangling bonds (if bulk termination is assumed) and has only two connecting bonds to the rest of the tip; dangling bonds are unsatisfied valences associated with atoms in the surface layer of a solid. If we assume bulk termination, it is immediately evident that tips pointing in the [111] direction are more stable, because then the front atom has three bonds to the rest of the tip.

The force is given by

$$F = -kz \tag{1.47}$$

where
  k is the spring constant
  z is the deflection of the cantilever

While scanning the tip across the surface, the force is kept constant. Then the vertical movement of the tip follows the surface profile and is recorded as the surface topography by the AFM. To form a map of the measured property relative to the X–Y position, these data are collected as the probe is scanned in a raster pattern (a set of pixels arranged in rows and columns) across the sample. The information is fed to a computer, which generates a map of topography and/or other properties of interest. Areas ranging from less than $100\,nm^2$ to as large as about $100\,\mu m^2$ are imaged.

1.20.5 Static and Dynamic Atomic Force Microscopy

In AFM, the force F_{ts} that acts between the tip and sample is used as the imaging signal. In the static mode of operation, the force translates into a deflection $q' = F_{ts}/k$ of the cantilever. Because the deflection of the cantilever should be significantly larger than the deformation of the tip and sample, restrictions on the useful range of k apply. In the static mode, the cantilever should be much softer than the bonds between the bulk atoms in tip and sample. Interatomic force constants in solids are in a range from 10 to about $100\,N\ m^{-1}$—in biological samples, they can be as small as $0.1\,N\ m^{-1}$. Force

constant is the ratio of the force acting to restrain the relative displacements of nuclei in a molecule to its deformation from equilibrium position. Thus, typical values for k in the static mode are 0.01–5 N m^{-1}. The eigenfrequency f_0 should be significantly higher than the desired detection bandwidth (the width of the range of frequencies between upper cutoff and lower cutoff frequencies), that is, if 10 lines per second are recorded during imaging a width of say 100 atoms, f_0 should be at least $10 \times 2 \times 100\,s^{-1} = 2\,kHz$ in order to prevent resonant excitation of the cantilever.

In the dynamic operation modes, the cantilever is deliberately vibrated. The cantilever is mounted on an actuator to allow the external excitation of an oscillation. There are two basic methods of dynamic operation: amplitude modulation (AM) and frequency modulation (FM). AM is the encoding of a wave by variation of its amplitude in accordance with an input signal. FM is a method of altering a waveform by changing the instantaneous frequency. In AM-AFM, the actuator is driven by a fixed amplitude A_{drive} at a fixed frequency f_{drive}, where f_{drive} is close to but different from f_0. When the tip approaches the sample, elastic and inelastic interactions cause a change in both the amplitude and the phase (relative to the driving signal) of the cantilever; phase is a particular point or any distinct time period in the time of a cycle, measured from some arbitrary zero and expressed as an angle. These changes are used as the feedback signal. The change in amplitude in AM mode does not occur instantaneously with a change in the tip–sample interaction, but on a time scale of $\tau_{AM} = 2Q/f_0$. With Q factors reaching 100,000 in vacuum, the AM mode is very slow. This problem was solved by introducing the FM mode, in which the change in the eigenfrequency occurs within single oscillation cycle on a time scale of $\tau_{FM} = 1/f_0$. Both AM and FM modes were initially meant to be "noncontact" modes, that is, the cantilever was far away from the surface and the net force between the front atom of the tip and the sample was clearly attractive. The AM mode was later used very successfully at a closer distance range in ambient conditions involving repulsive tip–sample interactions. Using the FM mode in vacuum improved the resolution dramatically. Finally atomic resolution was obtained.

In static atomic force microscopy, the imaging signal is given by the DC deflection of the cantilever, which is subject to $1/f$ noise. Pink noise or $1/f$ noise is a signal or process with a frequency spectrum such that the power spectral density is inversely proportional to the frequency. In dynamic atomic force microscopy, the low-frequency noise (noise that has a frequency between 20 and 100–150 Hz) is discriminated if the eigenfrequency f_0 is larger than the $1/f$ corner frequency (the frequency at which the $1/f$ noise spectral density equals the white noise, a random signal or process with a flat power spectral density). With a bandpass filter (an electronic device or circuit that allows signals between two specific frequencies to pass, but rejects or attenuates signals at other frequencies) with a center frequency (midpoint in the pass band) around f_0, only the white noise density is integrated across the bandwidth B of the bandpass filter.

1.20.6 Classification of Modes of Operation of AFM on the Basis of Contact

1.20.6.1 Contact Mode

In the contact-AFM mode, the tip makes soft "physical contact" with the surface of the sample. It either scans at a constant small height above the surface or under the conditions of a constant force. In the constant-height mode, the height of the tip is fixed, whereas in the constant-force mode, the deflection of the cantilever is fixed and the motion of the scanner in Z-direction is recorded.

For contact mode AFM imaging, it is necessary to have a cantilever that is soft enough to be deflected by very small forces and has a high enough resonant frequency to not be susceptible to vibrational instabilities. Silicon nitride (Si_3N_4) tips are used for contact mode. In these tips, there are several cantilevers with different geometries attached to each substrate, resulting in different spring constants. The spring constant is the constant of proportionality k, which appears in Hooke's law: $F = -kx$, where F is the applied force and x is the displacement from equilibrium. It has units of force per unit length and is a measure of how stiff the spring is.

To avoid problems caused by capillary forces (the forces involving molecular adhesion by which the surface of a liquid in a tube is either elevated or depressed, depending on the cohesiveness of the liquid molecules; they arise from intermolecular attractive forces between the liquid and surrounding solid surfaces), which are generated by a liquid contamination layer usually present on surfaces in air, the sample can be studied while immersed in a liquid. This procedure is especially beneficial for biological samples.

The advantages of contact mode are (a) high scan speeds, (b) atomic resolution is possible, and (c) easier scanning of rough samples with extreme changes in vertical topography.

The disadvantages of this mode are as follows: (a) Lateral forces can distort the image. (b) Capillary forces from a fluid layer can cause large forces normal to the tip sample interaction. (c) Combination of these forces reduces spatial resolution and can cause damage to soft samples.

1.20.6.2 Noncontact Mode

In this mode, the probe operates in the attractive force region and the tip–sample interaction is minimized. The use of noncontact mode allowed scanning without influencing the shape of the sample by the tip–sample forces. In most cases, the cantilever of choice for this mode is the one having high spring constant of 20–100 N m^{-1} so that it does not stick to the sample surface at small amplitudes. The tips mainly used for this mode are silicon probes.

An advantage of noncontact mode is that low force is exerted on the sample surface and no damage is caused to soft samples.

The disadvantages of noncontact mode are (i) lower lateral resolution, limited by tip–sample separation; (ii) slower scan speed to avoid contact with

fluid layer; and (iii) usually only applicable in extremely hydrophobic (repelling, tending not to combine with, or incapable of dissolving in water) samples with a minimal fluid layer.

1.20.6.3 Tapping Mode (Intermittent-Contact Mode)

In tapping-mode AFM the cantilever is oscillating close to its resonance frequency. An electronic feedback loop ensures that the oscillation amplitude remains constant, such that a constant tip–sample interaction is maintained during scanning. Forces that act between the sample and the tip will not only cause a change in the oscillation amplitude, but also a change in the resonant frequency (natural frequency of vibration determined by the physical parameters of a vibrating object; it is easy to make an object to vibrate at its resonant frequencies, but difficult to cause vibrations at other frequencies) and phase (the fraction of a wave cycle that has elapsed relative to an arbitrary point) of the cantilever. The amplitude is used for the feedback and the vertical adjustments of the piezoscanner are recorded as a height image. Simultaneously, the phase changes are presented in the phase image (topography). Phase imaging refers to recording the phase shift signal in intermittent-contact AFM. It is a powerful technique for producing contrast on heterogeneous samples. The phase shift can be thought of as a "delay" in the oscillation of the cantilever as it moves up and down in and out of contact with the sample. Phase of the cantilever oscillations, ϕ, is measured relative to the drive signal oscillations. Hence, phase imaging implies the monitoring of the phase lag between the signal driving the cantilever oscillations and the cantilever oscillation output signal. Therefore, changes in the phase lag reflect changes in the mechanical properties of the specimen surface.

Phase images compliment topography images by mapping the various regions of the sample surface each of which interact with the tip in a slightly or significantly different way. This difference is sometimes so subtle that it is barely noticeable in the topography image, but clearly visible in the contrast variations in the phase image.

Reverting to the mainstream discussion on the tapping mode, an advantage of the tapping mode is that there is almost no lateral force. Elimination of a large part of permanent shearing forces (internal forces in any material that are usually caused by external forces acting perpendicular to the material, or forces that have a component acting tangent to the material, leading to cutting and separating of the material through its cross section) create less damage to the sample surface, even with stiffer probes. Different components of the sample that exhibit different adhesive and mechanical properties will show a phase contrast and therefore even allow a compositional analysis. For a good phase contrast, larger tip forces are advantageous, while minimization of this force reduces the contact area and facilitates high-resolution imaging. So in applications it is necessary to choose the correct values matching the
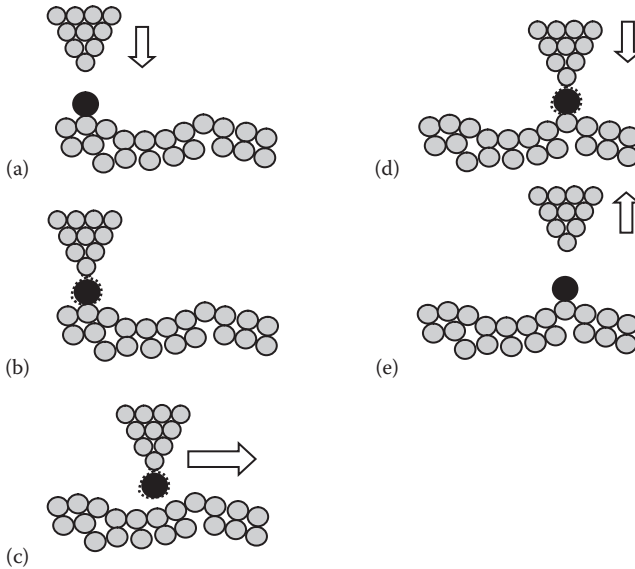
FIGURE 1.19
Manipulation of an atom with an AFM probe: (a) probe tip approaching the atom, (b) probe tip touching the atom, (c) probe tip transferring the atom to the new location, (d) probe tip placing the atom at the new position, and (e) withdrawal of the probe tip.

objectives. Silicon probes are used primarily for tapping-mode applications. Higher lateral resolution is achieved (1–5 nm).

A disadvantage of this mode is its slower scan speed than in contact mode.

The tip of the AFM is used for (i) measuring forces (and mechanical properties) at the nanoscale, (ii) for imaging, and (iii) as a nanoscale tool, that is, for bending, cutting, and extracting soft materials (such as polymers, DNA, and nanotubes), at the submicron scale under high-resolution image control. AFM is used as nanorobot for manipulating and controlling nano-objects, as shown in Figure 1.19.

1.20.7 Frequency-Modulation Atomic Force Microscopy

In frequency-modulation atomic force microscopy (FM-AFM) (Giessibl 2003), a cantilever with eigenfrequency f_0 and spring constant k works under controlled positive feedback (in which a portion of the output is combined in phase with the input), which ensures its oscillation with constant amplitude A, as shown in Figure 1.20. The deflection signal enters a bandpass filter, the output from which is divided into three branches: The first branch is shifted in phase, routed through an analog multiplier (device that produces an output voltage or current that is proportional to the product of two or more independent input voltages or currents), and fed back to the cantilever via an actuator, a mechanical device for moving or controlling
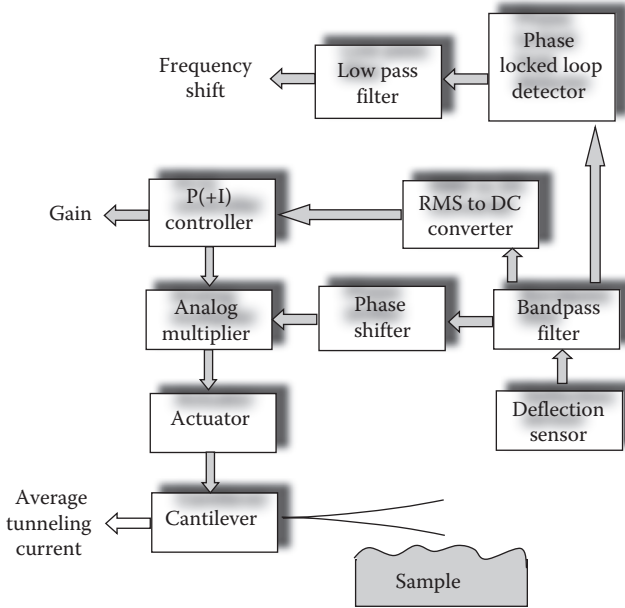
FIGURE 1.20
Frequency-modulation AFM feedback loop.

a mechanism or system. The second branch is used to compute the actual oscillation amplitude—this signal is applied to calculate a gain input g for the analog multiplier. The third branch feeds a frequency detector. In frequency detectors, frequency-modulated oscillations are first converted into amplitude-modulated oscillations, which are then detected by an amplitude detector. The frequency f is decided by the eigenfrequency f_0 of the cantilever and the phase shift ϕ between the mechanical excitation generated at the actuator and the deflection of the cantilever. If $\phi = \pi/2$, the loop oscillates at $f = f_0$.

Forces between tip and sample cause a change in $f = f_0 + \Delta f$. The eigenfrequency of a harmonic oscillator $= (k^*/m^*)^{0.5}/(2\pi)$, where k^* is the effective spring constant and m^* is the effective mass; a harmonic oscillator is a system that, when displaced from its equilibrium position, experiences a restoring force, F, proportional to the displacement, x. If the second derivative of the tip–sample potential $k_{ts} = \partial^2 V_{ts}/\partial z^2$ is constant for the whole range covered by the cantilever, $k^* = k + k_{ts}$. If $k_{ts} \ll k$, the oscillating square root is expanded as a Taylor series (series expansion of a function as an infinite sum of terms calculated from the values of its derivatives at a single point) and the shift in eigenfrequency is approximately

$$\Delta f = \left(\frac{k_{ts}}{2k} \right) f_0 \tag{1.48}$$

By measuring the frequency shift Δf, the tip–sample force gradient is determined.

1.20.8 Generic Calculation

The oscillation frequency is the main observable in FM-AFM, and it is useful to formulate a relation between frequency shift and the forces acting between tip and sample (Giessibl 2003). While the frequency can be calculated numerically, an analytic calculation helps in finding the functional relationships between operational parameters and the physical tip–sample forces. The motion of the cantilever (spring constant k, effective mass m^*) is described by a weakly disturbed harmonic oscillator. The deflection of the tip of the cantilever is $q'(t)$. It oscillates with an amplitude A at a distance $q(t)$ from a sample. The closest point to the sample is $q = d$ and $q(t) = q'(t) + d + A$. The Hamiltonian of the cantilever is

$$H = \frac{p^2}{2m^*} + \frac{kq'^2}{2} + V_{ts}(q) \tag{1.49}$$

where $p = m^* dq'/dt$. The unperturbed motion is represented by

$$q'(t) = A\cos(2\pi f_0 t) \tag{1.50}$$

and the frequency is

$$f_0 = \left(\frac{1}{2\pi} \right) \sqrt{\frac{k}{m^*}} \tag{1.51}$$

If the force gradient $k_{ts} = -\partial F_{ts}/\partial z$ is constant during the oscillation cycle, the frequency shift is calculated by the equation derived in the preceding section. However, in classic FM-AFM, k_{ts} varies by orders of magnitude during one oscillation cycle, and a perturbation approach (mathematical methods that are used to find an approximate solution to a problem that cannot be solved exactly, by starting from the exact solution of a related problem) is employed for the calculation of the frequency shift. This approach derives the magnitude of the higher harmonics and the constant deflection of the cantilever. The method involves the solution of Newton's equation of motion for the cantilever (effective mass μ^*, spring constant k):

$$\mu^* \left(\frac{d^2 q'}{dt^2} \right) = -kq' + F_{ts}(q') \tag{1.52}$$

The cantilever motion is assumed to be periodic and therefore is expressed as a Fourier series (an expansion of a periodic function in terms of an infinite sum of sines and cosines) with fundamental frequency f:

$$q'(t) = \sum_{m=0}^{\infty} a_m \cos(m2\pi ft) \tag{1.53}$$

Inserting this into Newton's equation yields

$$\sum_{m=0}^{\infty} a_m \left\{ -(m2\pi f)^2 \mu^* + k \right\} \cos(m2\pi ft) = F_{ts}(q') \tag{1.54}$$

Multiplication by $\cos(l2\pi ft)$ and integration from $t = 0$ to $t = 1/f$ gives

$$a_m \left\{ -(m2\pi f)^2 \mu^* + k \right\} \pi(1 + \delta_{m0}) = 2\pi f \int_0^{1/f} F_{ts}(q') \cos(m2\pi ft)\, dt \tag{1.55}$$

by applying the orthogonality of the angular functions

$$\int_0^{2\pi} \cos(mx)\cos(lx)dx = \pi \delta_{ml}(1 + \delta_{m0}) \tag{1.56}$$

Two functions f and g are called orthogonal if their inner product is zero for $f \neq g$. The *inner product* (or dot product, or scalar product) is an operation on two vectors that produces a scalar. For a weak perturbation, $q'(t) = A \cos(2\pi ft)$ with $f = f_0 + \Delta f$, $f_0 = (1/2\pi)\sqrt{(k/\mu^*)}$ and $|\Delta f| \ll f_0$. To first order, the frequency shift is given by

$$\Delta f = -\left(\frac{f_0^2}{kA} \right) \int_0^{1/f_0} F_{ts}(q') \cos(m2\pi f_0 t)dt = -\left(\frac{f_0}{kA^2} \right)(F_{ts}q') \tag{1.57}$$

which equals the result of the Hamilton–Jacobi method, a reformulation of classical mechanics and, thus, equivalent to other formulations such as Newton's laws of motion, Lagrangian mechanics, and Hamiltonian mechanics.

1.21 Outline and Organization of the Book

Sensor technology, a key technology in healthcare instrumentation, environmental monitoring, homeland security, aerospace and automotive sectors,

and industrial process control, has received a boost from new materials, and novel device structural designs and innovative concepts based on the foundations of nanoscience and nanotechnology. The development of fast and reliable nanosensors has been the focus of intensive research world-wide during the last two decades. Prime identifiable motives for progress in nanosensors have been the quest for high sensitivity, selectivity, resolution, repeatability, reliability, desirable hysteresis, and linearity properties in devices, and to develop subminiature sensors at lower effective costs, yielding high performance-cost ratios. Although probes for scanning and atomic force microscopes still dominate the commercial nanosensor arena, some new nanosensors, resulting from multidisciplinary research by high-tech companies and academic research laboratories, are beginning to make their way to the market place.

In this book, the progress in the field of nanosensors will be outlined, fundamental issues and challenges faced will be addressed and prospects for future development will be indicated. For convenience of study, the nanosensors will be systematically classified into physical (mechanical and acoustical, thermal and radiation, optical, magnetic) and chemical (atomic and molecular energies) categories. This classification scheme is convenient because a key characteristic of a sensor is conversion of energy from one form to another; hence, it is expedient to organize the classification according to the various forms of energy.

The biological domain overlaps both physical and chemical domains, and will be covered under these heads as an inter-mixed or interdisciplinary area because biosensors incorporate biological recognition elements with a physical or chemical sensor. Micro-/nano-cantilevers, for instance, overlap both nano-mechanical and nanobiosensors. In addition, nanobiosensors will be exclusively discussed in length in a separate chapter because of their enormous potential.

Chapter 1 will treat the nanosensor classification and fundamental terms. The properties of some important nanomaterials used in nanosensor fabrication will be outlined in Chapter 2. Nanotechnologies that are widely used in nanosensor fabrication will be briefly recapitulated in Chapter 3. Chapters 4 through 7 and Chapter 9 will describe the representative sensors belonging to sensor families characterized by the input signal to which they respond while Chapter 8 will be devoted to nanobiosensors. The book will conclude in Chapter 10 by summarizing the present overall nanosensor scenario and discussing future trends.

Thus, the proposed book will serve as a complete, definitive, and authoritative guide to nanosensors offering information on fabrication, properties, and operating mechanisms of nanosensors.

1.22 Discussion and Conclusions

This chapter introduced the definitions and concepts for understanding nanosensors and also presented the tools for viewing and chemically analyzing nanomaterials. It enabled us to learn the following: *what are nanosensors? How can they or their constituent parts be seen and their composition known?* Two important nanosensors, those for distance (STM) and force (AFM), were described.

Review Exercises

1.1 Define the following terms: (a) Natural Science, (b) physics, (c) chemistry, (d) biology, and (e) nanotechnology.

1.2 Which represents the diameter of a hydrogen atom: (a) 0.1, (b) 0.32, and (c) 0.04 nm?

1.3 What is the minimum length that the human eye can see?

1.4 What is a nanomaterial? Define 0D, 1D, and 2D nanomaterials. Give examples of each type.

1.5 Mention some unusual properties of matter that emerge at nanoscale.

1.6 Distinguish between sensors and transducers giving examples. Justify the statement, "All transducers contain a sensor and mostly, although not always, sensors will also be transducers."

1.7 List and describe the important parameters and characteristics of a sensor.

1.8 Mention two advantages of miniaturization of sensors. Explain and exemplify the statement "A miniaturized sensor can accomplish many tasks that a bulky device cannot perform."

1.9 Define a nanosensor. Give one example of a nanosensor that performs nanoscale measurements but does not have dimensions in the nano-range.

1.10 Give two examples of natural nanosensors.

1.11 Which of the following microscopes does not need a vacuum environment for its operation: (i) SEM, (ii) TEM, and (iii) AFM.

1.12 Explain the statement "SEMs are used as workhorses for routine work whereas TEMs are engaged in specialized tasks."

1.13 How does EDX complement the operation of SEM?

1.14 What is the advantage derived by using vacuum with AFM or STM?

1.15 Describe the contact, noncontact, and intermittent modes of operation of AFM, and point out their relative advantages and disadvantages.

1.16 Write the equation and draw the graph of Lennard–Jones potential for interaction between a pair of atoms.

1.17 IR absorption information is generally presented as a spectrum. What are the parameters plotted on the *X*- and *Y*-axes?

1.18 Can IR spectroscopy identify nonpolar molecules? What is Raman effect? Explain the principle of Raman spectroscopy.

1.19 Prepare a comparative chart of IR and UV–vis spectroscopy in terms of principle of operation, instrument construction, and fields of application.

1.20 Explain: (i) x-ray diffraction, (ii) XPS, and (iii) SIMS.

References

Booker, R. and E. Boysen. 2008. *Nanotechnology*. New Delhi, India: Wiley India Pvt. Ltd.

Bowker, M. and P. R. Davies (Eds.). 2010. *Scanning Tunneling Microscopy in Surface Science, Nanoscience and Catalysis*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KgaA.

Bréchignac, C., P. Houdy, and M. Lahmani (Eds.). 2007. *Nanomaterials and Nanochemistry*. Berlin, Germany: Springer-Verlag (EMRS), 753 pp.

Briggs, D. and J. T. Grant. 2006. *Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy*. Manchester, U.K.: SurfaceSpectra Ltd.

Brooks, K. (Ed.). 2006. *Chambers Science Factfinder*. Edinburgh, U.K.: Chambers Harrap Publishers Ltd.

Campbell, N. A. and J. B. Reece. 2004. *Biology*. San Francisco, CA: Benjamin Cummings, 312 pp.

Cao, G. 2004. *Nanostructures and Nanomaterials: Synthesis, Properties and Applications*. London, U.K.: Imperial College Press, 452 pp.

Chen, C. J. 2008. *Introduction to Scanning Tunneling Microscopy*. New York: Oxford University Press.

Curtis, L. J. 2003. *Atomic Structure and Lifetimes: A Conceptual Approach*. Cambridge, UK: Cambridge University Press, 282 pp.

Daintith, J. (Ed.). 2010. *A Dictionary of Physics*. New York: Oxford University Press, 624 pp.

Ferraro, J. R. 2003. *Introductory Raman Spectroscopy*. San Diego, CA: Academic Press, 434 pp.

Ghuo, Z. and L. Tan. 2009. *Fundamentals and Applications of Nanomaterials*. Norwood, MA: Artech House, 249 pp.

Giessibl, F. J. 2003. Advances in atomic force microscopy. *Reviews of Modern Physics* 75: 949–983.

Goldstein, J., D. E. Newbury, D. C. Joy, C. E. Lyman, P. Echlin, E. Lifshin, L. Sawyer, and J. R. Michael. 2003. *Scanning Electron Microscopy and X-Ray Microanaly*sis. New York: Springer Science + Business Media, Inc., 689 pp.

Hamaker, H. C. 1937. The London–van der Waals attraction between spherical particles. *Physica (Amsterdam)* 4: 1058–1072.

Hoffman Lab. 2010a. STM: More technical details, http://hoffman.physics.harvard.edu/research/STMtechnical.php

Hoffman Lab. 2010b. STM Measurement types, hoffman.physics.harvard.edu/research/STMmeas.php

Hollas, J. M. and M. J. Hollas. 2004. *Modern Spectroscopy*. New York: John Wiley & Sons, 480 pp.

Israelachvili, J. 1991. *Intermolecular and Surface Forces*, 2nd edn. London, U.K.: Academic Press.

Morris, V. J., A. R. Kirby, and A. P. Gunning. 2004. *Atomic Force Microscopy for Biologists*. London, U.K.: Imperial College Press, 324 pp.

Perkampus, H.-H. and H. Charlotte Grinter. 1992. *Uv–vis Spectroscopy and Its Applications (Springer Laboratory)*. Berlin, Germany: Springer-Verlag, 234 pp.

Smith, B. C. 1995. *Fundamentals of Fourier Transform Infrared Spectroscopy*. Boca Raton, FL: CRC Press, 224 pp.

Sze, S. M. (Ed.). 1994. *Semiconductor Sensors*. New York: John Wiley & Sons, Inc.

Tro, N. 2010. *Principles of Chemistry: A Molecular Approach*. Upper Saddle River, NJ: Pearson/Prentice Hall, 888 pp.

University of Wisconsin. 2007. ScanningTunneling Microscope (revised 1/9/07), http://www.hep.wisc.edu/~prepost/407/istm/istm.pdf

Wilson, J. 2004. *Sensor Technology Handbook*. London, U.K.: Newnes, 704 pp.

Zhu, Y., F. Xu, Q. Qin, W. Y. Fung, and W. Lu. 2009. Mechanical properties of vapor–liquid–solid synthesized silicon nanowires. *Nano Letters* 9(11): 3934.

2

Materials for Nanosensors

2.1 Introduction

When constructing a house, several materials are required, such as cement, concrete, bricks, tiles, iron and steel bars, flooring materials, and so on. Similarly, a large variety of materials is used to fabricate nanosensors. In fact, numerous! But some materials like nanoparticles (NPs), quantum dots (QDs), carbon nanotubes, inorganic nanowires, thin films, and nanoporous materials find widespread usage in many nanosensors. One repeatedly stumbles across their names while perusing nanosensor literature.

Low-dimensional nanometer-sized materials and systems have defined a new research area in condensed-matter physics (the field of physics dealing with the physical properties of condensed phases of matter such as solids and liquids) within the last 20 years. These will be described in this chapter. Apart from the aforesaid categories of materials, there exist innumerable materials of different types for fabricating nanosensors. These will be mentioned in specific examples.

Information presented in this chapter will help the reader to appreciate the roles and importance of nanomaterials for nanosensor fabrication.

2.2 Nanoparticles or Nanoscale Particles, and Importance of the Intermediate Regime between Atoms and Molecules, and Bulk Matter

The first term in the discussion of nanomaterials is the "nanoparticle." An NP is a particle of a metal, polymer, or oxide, smaller than 100 nm in diameter (Rotello 2003, Yang 2003, Schmid 2004, Hosokawa et al. 2007, Eftekhari 2008, Gubin 2009). It is essentially a zero-dimensional nanostructure.

Will a fiber thinner than 100 nm be considered a nanoparticle even if it is several micrometers long? No, it is a nanowire. This is essentially the case for carbon nanotubes which are definitely not "nano" in length, but they have a diameter in the order of 3 nm for a single-walled tube, and hence can be described as nanowires.

What is a one-dimensional nanostructure? A nanowire or nanorod, which is less than 100 nm in diameter but can be several micrometers long. Examples are nanotubes, nanocables, whiskers, fibers, etc.

What is a two-dimensional nanostructure? A thin film with thickness below 100 nm but may extend to hundreds of microns in other two dimensions.

Are nanoparticles visible? NPs, like viruses (small infectious agents that can replicate only inside the cells of other organisms), are invisible even through the best light microscope, because they are smaller than wavelengths of light (700 nm in the red to 400 nm in the violet). They can be seen only with higher-resolution instruments such as a scanning electron microscope.

Is a single sugar molecule having 1 nm size, a nanoparticle? No, because NPs are aggregates of atoms bridging the continuum between *small molecular clusters* of a few atoms and dimensions of 0.2–1 nm and *bulk solids* containing millions of atoms and having the properties of macroscopic bulk material. They have properties neither those of atoms (smallest unit into which an element can be divided and still retain the chemical properties of the element) and molecules (smallest division of a substance, element or compound, that still exhibits all the chemical properties of the substance; it is a group of similar or dissimilar atoms held together by chemical forces) or those of bulk materials (an assembly of solid particles that is large enough for the statistical average of any property to be independent of the number of particles) but somewhat intermediate between these properties.

In nanotechnology, not only NPs but one- and two-dimensional nanostructures are also included. The interesting properties displayed by all these nanostructures differ from atomic or molecular scale behavior as well as bulk characteristics. The fabrication, properties, and applications of these structures, *which is neither like atoms and molecules nor like bulk matter*, is what nanotechnology is all about.

It must be emphasized that nanotechnology is concerned with this *intermediate or in-between zone* between these two extreme limits. While bridging the gap between an atomic state and bulk phases of materials, nanoscale materials reveal novel physical and chemical properties which are completely different from those observed in either state of the materials. Size-dependent optical responses, exotic structural configurations, reaction catalysis (acceleration of a chemical reaction by the presence of material that is chemically unchanged at the end of the reaction), etc., are some novel properties exhibited by various classes of nanomaterials.

2.3 Classification of Nanoparticles on the Basis of Their Composition and Occurrence

How are the NPs broadly classified? There are two main classification schemes of NPs based on their composition and occurrence. NPs are classified

TABLE 2.1

Classification of Nanomaterials according to Composition

Sl. No.	Name of the Class	Examples
1	Metal-based	Metallic NPs, e.g., Au, Ag, etc.; metal oxides, such as zinc or titanium oxides
2	Carbon-based	Fullerenes, buckyballs, carbon nanotubes
3	Semiconductor-based	QDs, used in exploratory medicine or in the self-assembly of nanoelectronic structures
4	Polymer-based	Dendrimers (branched polymers)
5	Composite	Nanoclays, naturally occurring plate-like clay particles that strengthen or harden materials or make them flame-retardant. DNA molecules may be combined with various nanomaterials to make a nanosized biocomposite

TABLE 2.2

Classification of Nanomaterials according to Origin or Occurrence

Sl. No.	Name of the Class	Explanation	Remarks
1	Natural	Present in the environment as volcanic dust, lunar dust, mineral composites, etc.	May have irregular or regular shapes
2	Incidental (waste or anthropogenic particles)	Result from man-made industrial processes such as diesel exhaust, coal combustion, welding fumes, etc.	May have irregular or regular shapes
3	Engineered	Produced either by milling or lithographic etching of a large sample to obtained nanosized particles or by assembling smaller subunits through crystal growth or chemical synthesis to grow NPs of the desired size and configuration	Most often have regular shapes, such as tubes, spheres, rings, etc.

according to their chemical composition into five categories depending on whether they are based on metals, carbon, semiconductors, polymers, or composite materials (Table 2.1). From the viewpoint of their occurrence, they are subdivided into three classes, namely, natural, incidental (occurring as an unpredictable or minor accompaniment, as a subordinate or by chance), or engineered (Table 2.2).

2.4 Core/Shell-Structured Nanoparticles

What are structured nanoparticles and what is the advantage of structural configuration? Recently, core/shell NPs are finding widespread application (Sounderya

and Zhang 2008). Structurally, core/shell NPs are nanostructures that have the core made of a material coated with another material. These NPs have a size range of 20–200 nm. The necessity to shift to core/shell NPs arises from the quest for improvement in the properties of the base NPs. Taking into consideration the size of the NPs, the shell material is chosen such that the agglomeration of particles is prevented. This implies that the monodispersity (having the same size and shape) of the particles is improved.

Besides agglomeration prevention, does the core/shell structure improve any other property of the base nanoparticle? The core/shell structure also enhances the thermal and chemical stability of the NPs, improves their solubility, makes them less cytotoxic (poisonous to living cells) and allows conjugation (the state of being joined together) of other molecules to these particles. In some cases, the shell also prevents the oxidation of the core material.

When a core NP is coated with a polymeric layer or an inorganic layer like silica (SiO_2), synergistically emerged functions can be envisioned because the polymeric or inorganic layer would endow the hybrid structure with an additional function/property on top of the function/property of the core.

2.4.1 Inorganic Core/Shell Nanoparticles

What is the composition of inorganic core/shell nanoparticles, structurewise? The core or the shell or both are made of inorganic materials.

What are metallic core/shell nanoparticles? In metallic core/shell NPs, the core is a metal/metal oxide or silica while the shell is silica or a metal/metal oxide. The most widely used core/shell nanocomposites are gold or silver core with silica shell. The gold/silica NPs are used in optical sensing, and the thickness of the silica coat alters the optical property of gold NPs.

What are semiconductor core/shell nanoparticles? In semiconductor NPs, the core is made of semiconductor material, semiconductor alloy, or metal oxide with shell made of semiconductor material, metal oxide, or an inorganic material like silica.

What are the specifics of these structures? These structures can be binary (consisting of two parts or components) with a core and shell or a ternary (having three elements, parts, or divisions) structure with a core and two shells. The most common binary structure that are well known by the name QDs are an alloy of group III and group V metals or group IV and group VI metals, namely, CdSe/CdS, CdSe/ZnS, ZnSe/ZnS, CdTe/CdS, etc; the shell thickness determines the emission range of these particles. They fall under the category of binary NPs (Li et al. 2009).

What are Lanthanide nanoparticles? Lanthanide (rare earth: any element of the lanthanide series [atomic numbers 57–71]) NPs have a core which contains one or more lanthanide group elements surrounded by a shell made of inorganic material like silica or a lanthanide material. Aqueous colloids of Rhabdophane ((Ce,Y,La,Di)(PO$_4$)·H$_2$O), brown, pinkish, or yellowish-white mineral consisting of a hydrated phosphate of cerium, yttrium, and rare

earths are one of these categories that show a green luminescence. These are Ce, Tb-doped core particles with an $LnPO_4–xH_2O$ shell. These particles can be further coated with silica to enhance their luminescent properties. They have potential applications in electronics and bioimaging.

2.4.2 Organic–Inorganic Hybrid Core/Shell Nanoparticles

What are the main subdivisions of hybrid nanoparticles? These include: (i) Organic core and inorganic shell NPs: polyethylene $((–CH_2–CH_2–)_n)$/silver (Ag), poly-lactide (PLA: $C_3H_6O_3$)/gold (Au). (ii) Inorganic core and organic shell NPs: SiO_2/PAPBA(poly(3-aminophenylboronic acid), Ag_2S/PVA(polyvinylalcohol), CuS/PVA, Ag_2S/PANI(polyaniline), and TiO_2/cellulose; 3-aminobenzenebo-ronic acid monohydrate $= C_6H_8BNO_2 \cdot H_2O$; PVA $= (C_2H_4O)_x$; aniline $= C_6H_5NH_2$; cellulose $= (C_6H_{10}O_5)_n$ (iii) Polymeric core/shell NPs: polymethylmethacrylate (PMMA:$(C_5O_2H_8)_n$) coated antimony trioxide (Sb_2O_3) compounded with poly-vinylchloride (PVC:$CH_2 = CHCl$)/antimony trioxide composites. The interaction between PMMA and the PVC along with antimony trioxide enhances toughness and strength of PVC.

2.5 Shape Dependence of Properties at Nanoscale

Engineered nanomaterials with identical chemical composition have a variety of shapes such as spheres, tubes, fibers, rings, and planes. *Do these particles of different shapes exhibit dissimilar properties?* Yes, every one of these shapes may have different physical properties, because the pattern of molecular bonds differ even though they are composed of the same atoms, for example, the properties of fullerenes (molecules composed entirely of carbon; buck-minsterfullerene $= C_{60}$) differ from those of its other two allotropic forms of carbon, namely, diamond and graphite.

In diamond, every C-atom is sp^3 hybridized and covalently bound to four other C-atoms. The binding arms of these C-atoms point to the corners of a tetrahedron. There is an angle of 109° between each arm resulting in the unique structure of diamond. In graphite, every C-atom is sp^2 hybridized and covalently bound to three other C-atoms. Thus a plane of continuous hexagons results with feeble van der Waal's forces acting between the hexagons. In fullerenes, a spherical network exists. In contrast to diamond and graphite which are made of expanded three-dimensional structures, fullerenes form closed molecular systems with a sp^2 hybridized C-atom as a common building block. They contain pentagons and hexagons; the pentagons are responsible for the bend and the football structure of the fullerenes. As a result, they dissolve in various solvents which makes their chemical manipulation easier. The available fullerenes are C_{60}, C_{70}, C_{76}, C_{78}, C_{84}, C_{90}, C_{94}, and C_{96}.

2.6 Dependence of Properties of Nanoparticles on Particle Size

Does the nanoparticle shape only matters? At the NP scale, both size and shape of the particle are important. One reason of size dependence is increase in surface area with diminishing particle size. Surface area plays a definitive role here because most chemical reactions involving solids happen at the surfaces, where chemical bonds are incomplete. Thus, collections of nanoscale particles (NSPs) with their enormous surface areas are exceptionally reactive (unless a coating is applied), because more than a third of their chemical bonds are at their surfaces. For example, NPs of silver have been found to be an effective bactericide (a substance that kills bacteria), inspiring several companies to design reusable water-purification filters using nanoscale silver fibers.

Example 2.1

A simple thought experiment reveals why NPs have such high surface area per unit volume. A solid cube of a material 1 cm on a side—about the size of a sugar cube—has 6 cm^2 of surface area (Figure 2.1). But if that volume of 1 cm^3 were filled with cubes 10 nm on a side, there would be $(10^{-2} \times 10^{-2} \times 10^{-2})$ m^3/$(10 \times 10^{-9} \times 10 \times 10^{-9} \times 10 \times 10^{-9})$ m$^3 = 10^{18}$ cubes of 10 nm-size cubes (volume = 10 nm × 10 nm × 10 nm), each one of which has a surface area of 6×10 nm × 10 nm = $6 \times 10 \times 10^{-9} \times 10 \times 10^{-9} = 6 \times 10^{-16}$ m^2, giving a total surface area = $6 \times 10^{-16} \times 10^{18}$ m$^2 = 6 \times 10^2$ m$^2 = 6 \times 10^2 \times 10^2 \times 10^2$ cm$^2 = 6 \times 10^6$ cm^2. Hence, the surface area increases by a factor of $(6 \times 10^6$ cm$^2)/6$ cm$^2 = 1 \times 10^6$.

2.7 Surface Energy of a Solid

How is surface energy of a solid defined? It is the sum total of energies of all the atoms or molecules present at the surface of a solid.
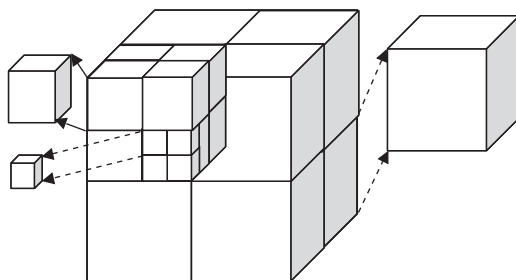
FIGURE 2.1
Subdivision of a cube into smaller cubes.

A solid is formed when individual atoms come closer and distribute themselves to satisfy their bonds on all the four sides. In this way, the total energy of atoms decreases to reach a more stable state. But this condition of satisfied bonds does not hold for the surface atoms because their bonds toward the surface side are left dangling. Thus surface atoms have more energy as compared to subsurface atoms. This extra energy will be released and the surface will acquire a more stable state when the surface bonds are satisfied, for example, by the attachment of adsorbed atoms. The excess energy of surface atoms above those present in a bulk solid is collectively called the *surface energy of the solid*.

Surface energy is defined as the energy required to create a unit area of new surface.

After noting that the surface atoms of a solid possess more energy than those in the bulk solid, let us enquire about the unbalanced interatomic bond forces near the surface. *What is the effect of these forces on bond lengths*? Due to the dangling bonds at the surface, the surface atoms experience an inward directed force toward the bulk solid resulting in shortening of the bond lengths of surface atoms. Thus there is a decrease in bond lengths and lattice constants at the surfaces of solids.

2.8 Metallic Nanoparticles and Plasmons

Why have metallic nanoparticles aroused interest? NPs of gold, silver, and copper have received attention due to their attractive electronic and optical properties. These particles have the characteristic feature of displaying beautiful bright colors due to a property called surface plasmon band (SPB).

Is SPB unique to nanoparticles? This phenomenon is not unique to NPs but is also observed on the surfaces of bulk metals where it is called surface plasmon resonance (SPR). *What is the confusion in terminology regarding SPR and SPB?* There is a little confusion in terminology because *prima facie* it appears that SPB is not a resonance phenomenon but SPR is a resonance phenomenon. In reality, both are resonance phenomena, and SPB is sometimes referred to as *localized surface plasmon resonance* (LSPR) to distinguish it from SPR (Hutter and Fendler 2004, Dahlin et al. 2006).

What is a plasmon and how is this term coined? In general, a *plasmon* is a quantum of plasma oscillations; a plasma is a quasineutral gas of charged and neutral particles which exhibits collective behavior. The coinage of the term "plasmon" is similar to calling a quantum of light a "photon" or a quantum of lattice vibrations by the name "phonon." The plasmon is the short form of *plasmon polariton*.

Is plasmon a quasiparticle? Yes, it is a quasiparticle, a long-lived single particle excitation in which the excitations of individual particles are modified by

their interactions with surrounding medium. Excitation is promotion to a higher energy quantum state.

What is plasmonics? Study of optical phenomena in relation to electromagnetic response of metals is referred to as *plasmonics*.

The obvious question is: *Where is the plasma found in a metal?* A metal contains a sea of free electrons and positive ion cores. This ensemble of *free electron gas* and positive ions in a metal constitutes the plasma. Light consists of electromagnetic waves. The response of free electrons of the plasma to electric field variations in light waves generates plasma oscillations and hence plasmons.

How to visualize a plasma oscillation? Let us think of a metallic cube placed in an external electric field whose direction is toward the right hand. Under the influence of this electric field, electrons move toward the left-hand side thereby uncovering positive ions on the right-hand side. This continues until they cancel the field inside the metal. Now if the electric field is switched off, the electrons move to the right-hand side, repelled by each other and attracted to the positive ions left bare on the right side. Thus the electrons oscillate back and forth at the plasma frequency (the frequency at which plasma oscillations occur; clearly, there is a different plasma frequency for each species) until the energy is dissipated in some kind of resistance or damping. Plasmons are a quantization of this kind of oscillations.

What role is played by plasmons in the optical properties of metals? Plasmons play a key role in the optical properties of metals. Metals have a shiny appearance due to reflection of light from their surfaces back to the eye. The reason for this reflection is associated with the electron cloud surrounding the metals. Because the electron cloud in the metals screens the electric field of the light, photons having frequency less than the plasma frequency are reflected (Figure 2.2).

Light of frequency higher than the plasma frequency is transmitted, because the electrons cannot respond fast enough to screen it. Generally for metals, the plasma frequency is in the ultraviolet region (7.5×10^{14}–3×10^{16} Hz; 10–400 nm), making them shiny or reflective in the visible portion of spectrum (wavelengths from approximately 400–720 nm). But metals like copper and gold have electronic interband transitions in the visible range. Hence specific light energies (colors) are absorbed enabling the appearance of their distinct color. In semiconductors, the plasma frequency of valence band electrons is usually in the deep ultraviolet (DUV: <300 nm), so that they are also reflective.

2.8.1 Surface Plasmon Resonance on Bulk Metals

Surface plasmons are those plasmons that are confined to surfaces and interact strongly with light producing a polariton. They occur at the interface of a vacuum or material with a positive relative permittivity and a negative relative permittivity (usually a metal or doped dielectric).

Consider the excitation of a piece of bulk metal by electromagnetic waves (Figure 2.2b). Evanescent waves are produced. Evanescent waves (plasmons)
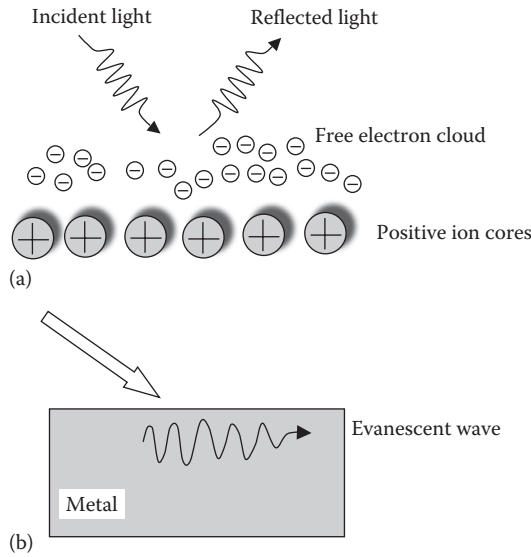
FIGURE 2.2
(a) Reflection of light by electron cloud in metals at lower frequencies than plasma frequency.
(b) Production of evanescent waves by excitation of bulk metal.

are waves that are decaying in amplitude having crossed into a medium attenuating them.

Surface plasmon waves are electromagnetic waves propagating at optical frequencies on an interface between a metal, typically gold or silver, and a dielectric. They are evanescent waves: their field intensity is concentrated in a very thin layer (few tens of nanometers) across the interface. They are characterized by a propagation constant k_{sp}. The propagation or wave vector of a wave is a vector quantity that defines the magnitude and direction of the wave; its magnitude is $2\pi/\lambda$. For a flat interface, the propagation vector of surface plasmon waves is given by

$$k_{sp} = k_0 \sqrt{\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d}} \qquad (2.1)$$

where
  ε_m and ε_d are the dielectric constant of the metal and the dielectric, respectively
  k_0 is the momentum of light in free space at the same frequency

According to Equation 2.1, the propagation constant of surface plasmons is strongly dependent on the variations of permittivity on the interface. Therefore, SPR is a *surface-sensitive optical technique* based on changes in refractive index at a metal–dielectric interface.
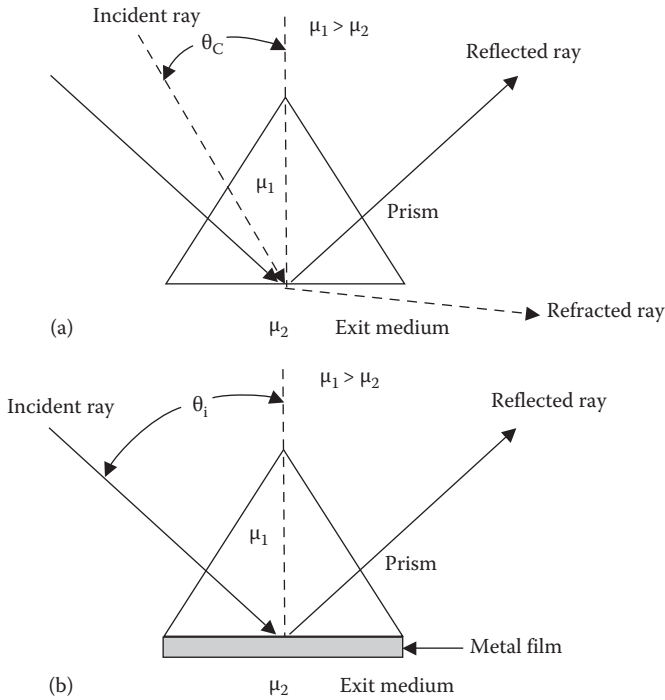
FIGURE 2.3
(a) Total internal reflection at an interface. (b) Use of a prism coupling configuration for SPR
(ICX Nomadics: Overview of surface plasmon resonance).

Surface plasmon resonance sensors are usually constructed by using prism
coupling of incident light onto an optical substrate that is coated with a semi-
transparent (partially transparent: not perfectly or completely transparent)
noble metal under conditions of total internal reflection (complete reflection
of a ray of light at the boundary between two transparent media, which
occurs if the angle of incidence is greater than a certain limiting angle,
called the critical angle). Therefore, a condition of total internal reflection
(Figure 2.3a) must exist at the interface. Total internal reflection will exist for
incident angles greater than critical angle θ_c; at this angle some of the light
is refracted across the interface. At the point of reflection at the interface,
an evanescent field (standing wave) penetrates the exit medium to a depth
in the order of ¼ of the incident light wavelength. A standing wave, also
known as a stationary wave, is a wave that remains in a constant position;
standing waves are produced whenever two waves of identical frequency
interfere with one another while traveling in opposite directions along the
same medium.

μ_1 and μ_2 are the refractive indices of the prism and exit medium, respec-
tively. The refracted beam at the critical angle is shown as a directed dotted

line. All light is reflected at incident angles $>\theta_c$. If a semitransparent noble metal film is placed at the interface (Figure 2.3b), then under conditions of total internal reflection, SPR occurs. This is commonly known as the *Kretschmann configuration*.

The condition of *surface plasmon resonance* will occur when the following conditions are satisfied. The incident wave vector (a vector which helps describe a wave and whose direction is the direction of wave propagation) is given by the equation (ICX Nomadics: Overview of surface plasmon resonance)

$$k_i = \left(\frac{2\pi}{\lambda}\right) \mu \sin\theta_i \tag{2.2}$$

where

k_i is a component of the incident light wave vector parallel to the prism interface
θ_i is the incident light angle
λ is the wavelength of the incident light
μ is the refractive index of the prism material

The wave vector of the plasmon mode is described by the equation

$$k_{sp} = \left(\frac{2\pi}{\lambda}\right) \sqrt{\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d}} \tag{2.3}$$

where

k_{sp} is the surface plasmon wave vector
ε_m and ε_d are the relative permittivity constants of the metal film and the dielectric exit medium, respectively

SPR takes place during the equality of wave vectors

$$k_i = k_{sp} \tag{2.4}$$

How do evanescent waves interact with incoming waves? The evanescent waves present at the surface of the metal interact with the incoming electromagnetic waves causing a perturbation of the reflected signal. The intensity of the reflected light will decrease at SPR thereby leading to a well-defined minimum in the reflectance intensity.

How is resonance observed? At a given wavelength, extinction of the reflected signal is observed which indicates the occurrence of resonance. Thus resonance is observed through extinction of the signal; extinction is the condition of being extinguished.

If the incident angle is fixed and polychromatic light (described by many different frequencies) is reflected from the surface, then light will be adsorbed by the resonance at particular wavelengths. This will give rise to a typical plasmon resonance minimum in the reflectance spectrum.
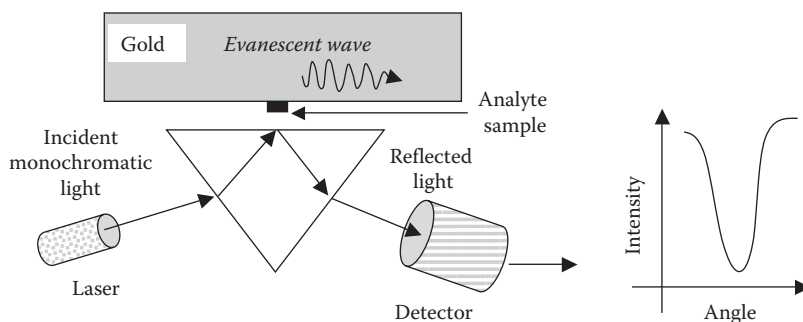
FIGURE 2.4
Schematic layout of experimental setup for conducting SPR.

If monochromatic light (light of one color) is reflected from the surface over a range of incident angles, then a similar reflectance minimum is observed with respect to the angle of incidence. The reflectance minimum that is produced from plasmon resonance is caused by the phase difference (the time interval or phase angle by which one wave leads or lags another) of the surface mode with respect to the incident photon field. Below the resonance, the phase difference is 0° and it approaches 180° above the resonance. Therefore, the photons that are reflected from the metal–prism interface undergo destructive interference (out-of-phase superposition yielding zero intensity) with the photons emitted by the excited plasmons that are 180° out of phase immediately above the maximum resonance. As a consequence, the characteristic reflectance minimum is produced.

How is an SPR experiment performed in the laboratory? A practical SPR set up consists of a laser source to illuminate the metal surface (Figure 2.4). This illumination is done either at a fixed angle and variable wavelength or at variable angle and fixed wavelength. The reflected signal is collected and measured by a detector.

When resonance takes place, the reflectance is minimum. Therefore, the situation of minimum reflectance is ascertained and the corresponding wavelength or angle is recorded.

For sensor applications, it is this change in the refractive index of the dielectric exit medium at the metal surface that is of interest. Therefore, if the refractive index of the prism is constant, then a change in the resonance condition may be linked to changes in the refractive index of the exit medium. In this way, it is possible to monitor the accumulation of films on the metal surface in order to measure binding of molecules to a surface that has been coated with an affinity (a natural liking for or attraction: the force attracting atoms to each other and binding them together in a molecule) ligand (an ion, a molecule, or a molecular group that binds to another chemical entity to form a larger complex). From these considerations, SPR configuration may be regarded as a *surface-sensitive refractometer* (instrument for measuring the
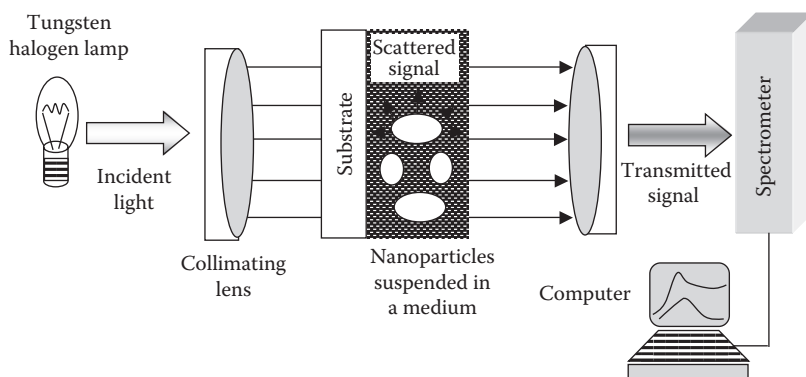
FIGURE 2.5
Schematic layout of experimental setup for SPB experiment.

refractive index of a substance) in which the sensitivity depth is defined by the penetration depth of the evanescent field (~200 nm).

2.8.2 Surface Plasmon Band Phenomenon in Metal Nanoparticles

Here the oscillation of three-dimensional confined free electron gas in metal NPs takes place in resonance with the incident optical signal. The resulting absorption and scattering phenomena (in which the direction, frequency, or polarization of the wave is changed when the wave encounters discontinuities in the medium, or interacts with the material at the atomic or molecular level) are known as SPB.

What is the main dissimilarity between SPB and SPR? The strikingly distinguishing feature of SPB with respect to SPR is the simplicity of the SPB instrument in which a laser source is neither mandatory nor is a complex detector necessary (Figure 2.5). The phenomenon is easily observed by the unaided eye. A UV-visible spectrometer is able to provide the desired information. Thus the instrument is portable besides being low cost and user-friendly. Naturally, therefore, the SPB instrument has received favorable attention of analytical scientists for carrying out their experiments.

Table 2.3 brings out a comparison between SPR and SPB.

TABLE 2.3

Comparison of Surface Plasmon Resonance and Surface Plasmon Band

Sl. No.	Surface Plasmon Resonance	Surface Plasmon Band
1	Takes place in bulk metals	Occurs in metallic NPs
2	It is a resonance phenomenon	It is a resonance phenomenon
3	Requires elaborate instrument setup including laser source and photodetector assembly	Needs simpler instrumental setup
4	Costly instrument	Relatively cheap instrument

2.9 Optical Properties of Bulk Metals and Metallic Nanoparticles

2.9.1 Light Absorption by Bulk Metals and Metallic Nanoparticles

What models are needed to explain light absorption by bulk metals and metallic nanoparticles? The absorption of light by bulk metals and NPs is interpreted by a model for describing waves along with a separate model for particles. The wave model is needed for electromagnetic waves and the particle model for electrons.

What are the suitable available models for this description? The appropriate wave model is Maxwell electromagnetic wave theory and particle model is Drude's theory of free electrons.

The equation of electromagnetic waves of frequency ω propagating in a metal of dielectric constant $\varepsilon(\omega)$ and conductivity $\sigma(\omega)$ is (Moores and Floch 2009):

$$\nabla^2 \mathbf{E} = -\mu_0 \varepsilon_0 \omega^2 \left\{ \varepsilon_\infty + \frac{i\sigma(\omega)}{\omega\varepsilon_0} \right\} \mathbf{E} \qquad (2.5)$$

where
  $\mathbf{E}$ is the electric field
  μ_0, ε_0 are respectively the permeability and permittivity of free space

This equation has essentially the same form as that of a wave traveling in free space:

$$\nabla^2 \mathbf{E} = -\mu_0 \varepsilon_0 \omega^2 \mathbf{E} \qquad (2.6)$$

Equation 2.6 becomes Equation 2.5 by inserting the frequency-dependent dielectric constant of the metal in the presence of the wave in this equation; $\varepsilon(\omega)$ is written as

$$\varepsilon(\omega) = \varepsilon_\infty + \frac{i\sigma(\omega)}{\omega\varepsilon_0} \qquad (2.7)$$

Implicit here is the fact that the dielectric constant of the metal is affected by the incident wave, that is, the incident wave alters the behavior of the metal. The influence of the wave on the metal is thus accounted for. Another important fact to be understood from Equation 2.7 is that $\varepsilon(\omega)$ is a complex quantity containing real and imaginary components signifying that the incident wave and the dielectric response are not necessarily in phase.

Whereas the free-space problem of Equation 2.6 is easy to solve, solution of Equation 2.5 is difficult. Although experimental approaches have been

followed, theoretical models were proposed. Drude's theory is used for this purpose. In this theory, the loosely bound conduction electrons are considered to be free and independent. They are treated classically, free to move within the crystal but colliding with each other and with the ion cores, which are assumed to be immobile. The motion of an electron cloud is represented by the sum total of the motions of individual electrons. All the electrons move in phase so that inter-electron coupling is maximized.

How do we write the equation of motion of an electron? For a single electron, the equation of motion is written by equating the force acting on the electronic charge q in the applied electric field E, that is, qE with the sum of the forces responsible for the motion of the electron and opposition to this motion by damping caused by various factors like inelastic collisions (collisions between two particles in which part of their kinetic energy is transformed to another form of energy) of free electrons with metal ions, electron–phonon coupling, and scattering of free electrons at impurities and defects in the metallic structure. The equation of electron motion is

$$m_e \left(\frac{dv}{dt} \right) + m_e \Gamma v = qE \tag{2.8}$$

where
  m_e is the effective mass of the electron (a parameter with the dimensions of mass that is assigned to electrons in a solid; in the presence of an external electromagnetic field the electrons behave in many respects as if they were free, but with a mass equal to this parameter instead of the true mass)
  v is its velocity
  Γ is the damping constant

It may be noted that the force applied by the magnetic field component of electromagnetic wave has been neglected, being very small as compared to the force due to electric field component of the same. *How is this justified?* Justification for this approximation is the very small velocity of electrons in comparison to the velocity of light.

For a sinusoidal electric field with maximum field value E_0,

$$E = E_0 \exp(-i\omega t) \tag{2.9}$$

the electron velocity will also vary sinusoidally as

$$v = v_0 \exp(-i\omega t) \tag{2.10}$$

where v_0 is the peak velocity. Hence, Equation 2.8 takes the form

$$(-i\omega m_e + m_e \Gamma)v_0 \exp(-i\omega t) = qE_0 \exp(-i\omega t) \tag{2.11}$$

giving

$$v_0 = \frac{qE_0}{m_e\Gamma - i\omega m_e} \tag{2.12}$$

The resulting electric current is

$$j = qnv \tag{2.13}$$

where n is the number of electrons per unit volume. In terms of the peak value j_0, the electric current is expressed as

$$j = j_0 \exp(-i\omega t) \tag{2.14}$$

where

$$j_0 = qnv_0 = \frac{q^2 n \mathbf{E}_0}{m_e\Gamma - i\omega m_e} = \sigma(\omega)\mathbf{E}_0 \tag{2.15}$$

Here $\sigma(\omega)$ is the conductivity of the metal and Equation 2.15 expresses the Joule's law. Equations 2.7 and 2.15 are combined together to relate the dielectric constant $\varepsilon(\omega)$ in terms of known constants $\varepsilon(\infty)$, n, q, m_e, ε_0, frequency ω, and damping constant Γ as (Moores and Floch 2009):

$$\varepsilon(\omega) = \varepsilon_\infty - \frac{\omega_P^2}{\omega^2 + i\omega\Gamma} = \varepsilon_\infty - \frac{\omega_P^2}{\omega^2 + \Gamma^2} + \frac{i\omega_P^2\Gamma}{\omega(\omega^2 + \Gamma^2)} \tag{2.16}$$

where, for simplification, we have put

$$\omega_P^2 = \frac{nq^2}{\varepsilon_0 m_e} \tag{2.17}$$

How is Γ determined? Γ is determined experimentally.

Equation 2.16 is valid for bulk metals. *What are the special considerations that must be taken into account for nanoparticles?* For NPs, the following conditions hold: (i) Current density ρ in the metal NPs is zero because electron density is not uniform in small particles. (ii) Because of the infinitesimally small size of the NPs relative to the wavelength of incident electromagnetic waves, at any given instant of time, all the electrons confined in an NP see and hence experience the same electric field. As a result, the electric field acting on the electrons in an NP is position-independent. Therefore, all these electrons behave identically and in harmony in response to the electromagnetic waves. Under the influence of the electric field, the electron clouds in the NP are displaced from their respective positions. In this way, charges are created on the surfaces of NPs through imbalancing of net charges in the particles. These surface charges
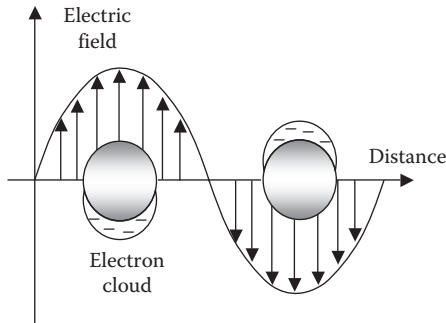
FIGURE 2.6
Electron cloud displacements in spherical metallic NPs with respect to nuclei, produced by an impinging electromagnetic wave. (After Moores, A. and Floch, P. L., The metal nanoparticle plasmon band as a powerful tool for chemo- and biosensing, In: *Biosensing Using Nanomaterials*, Merkoçi, A. (Ed.), John Wiley & Sons, Inc., New Jersey, 137, 2009.)

are negative at locations where there are higher concentrations of electrons (Figure 2.6). Conversely, the surface charges are positive where lesser concentrations of electrons exist. Hence, the surfaces of NPs play important roles in this phenomenon. So, it is natural to call it by the name "surface plasmon band."

Furthermore, it must be noted that the electrons act in coordination or unison, that is, collectively. All the electrons present in an NP are undergoing collective motion. The collective oscillations of electrons in response to the incident waves are known as plasmon polaritons or plasmons in contrast to the free plasmons in bulk metals. This explains the use of the word "plasmon" in this phenomenon.

As explained earlier, there is a dipolar charge separation in the NPs so that the NPs act as dipoles (a pair of electric charges, of equal magnitude but of opposite sign, separated by a small distance). *What is the effect of these dipolar forces?* Dipolar forces try to oppose the effects of incident waves and restore the electron clouds to their original positions. *How is influence of dipolar forces observed?* This is observed as a restitution or damping effect on the electron oscillations. Accordingly, Equation 2.8 is modified for the NP motion as

$$m_e \left(\frac{dv}{dt} \right) + m_e \Gamma v + Kx = qE \qquad (2.18)$$

where
  K is the restoring force constant
  x is the position of an electron in the electron cloud

In terms of x, Equation 2.14 is recast as

$$m_e \left(\frac{d^2 x}{dt} \right) + m_e \Gamma \left(\frac{dx}{dt} \right) + Kx = qE \qquad (2.19)$$

Thus the motion of an electron in the electron cloud of an NP is analogous to that of a damped harmonic oscillator, that is, a classical mechanical oscillator. A harmonic oscillator is a system which, when displaced from its equilibrium position, experiences a restoring force proportional to the displacement from the mean position; damping is any effect that tends to reduce its amplitude of oscillations.

Looking at the electric field seen by the NP, it is the one applied externally but altered by the polarizability (the electric dipole moment induced in a system, such as an atom or molecule, by an electric field of unit strength) of the medium of the NP. Applying the boundary condition in a spherical particle, the internal field $\mathbf{E}_i$ in an NP surrounded by vacuum is expressed as

$$\mathbf{E}_i = \mathbf{E}_0 \left\{ \frac{3}{\varepsilon(\omega) + 2} \right\} \tag{2.20}$$

Equation 2.20 gives the condition of resonance which occurs when $\mathbf{E}_i$ is maximum, that is, when $|\varepsilon(\omega) + 2|$ is minimum. In other words, this takes place when

$$\left| \varepsilon_1(\omega) + 2 \right|^2 + \left| \varepsilon_2(\omega) \right|^2 = \text{minimum} \tag{2.21}$$

because

$$\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega) \tag{2.22}$$

Thus for resonance frequency ω_M,

$$\varepsilon_1(\omega_M) = -2 \tag{2.23}$$

Considering $\varepsilon_\infty = 1$ and $\Gamma \ll \omega$, it readily follows from Equations 2.16 and 2.23 for the electron motion in an NP, described by Equation 2.19 that

$$\varepsilon_1(\omega_M) = 1 - \frac{\omega_P^2}{\omega_M^2} = -2 \tag{2.24}$$

from which we have

$$\omega_M = \frac{\omega_P}{\sqrt{3}} \tag{2.25}$$

Substituting for ω_P from Equation 2.17 we get

$$\omega_M = (q) \sqrt{\left(\frac{n}{3\varepsilon_0 m_e} \right)} \tag{2.26}$$

To summarize, the electron cloud in the NP oscillates in the incident light. Surface charges are produced at the edges of the NP leading to the formation of dipoles and thereby providing the restoring force. Resonance takes place at the frequency ω_M given by Equation 2.25.

Was the aforementioned treatment oversimplified for nanoparticles? Yes, the treatment of the phenomenon presented was overly simplified because only one isolated NP in vacuum was considered and several postulates were made including $\varepsilon_\infty = 1$ and $\Gamma \ll \omega$. Nonetheless, the physical picture of the phenomenon occurring in the NP could be clarified. Three forces, namely, those due to the electric field, restoring and damping effects acted on the NP.

Does an example of exploitation of metallic nanoparticle effects occur in history? Yes. The ancient Romans exploited the size dependence of the plasmon resonance frequency of gold at the nanoscale when they added gold salts during glass making to obtain red-colored glass due to the presence of gold nanocrystals. The remarkable Lycurgus cup in the British Museum is an example of this technique.

2.9.2 Light Scattering by Nanoparticles

What is the main idea of this approach? Here we note that under photoexcitation (the process of exciting the atoms or molecules of a substance by the absorption of radiant energy), the electrons in the electron cloud of the NP undergo acceleration. Since *an accelerated charge radiates energy in space in the form of electromagnetic waves,* the accelerated electrons themselves emit electromagnetic waves in all directions. Interaction takes place between the emitted waves and incoming waves. The emitted waves scatter the energy of the incoming unidirectional wave. Scattering is a general physical process where light is forced to deviate from a straight trajectory by one or more localized nonuniformities in the medium through which it passes. This also includes deviation of reflected radiation from the angle predicted by the law of reflection. Reflections that undergo scattering are often called diffuse reflections.

At the resonance condition, the scattering of energy is maximum. Hence, most of the incident waves are scattered and not transmitted. Therefore, the transmitted signal reaches a minimum. Thus the absorbance spectrum (a spectrum of radiant energy whose intensity at each wavelength is a measure of the amount of energy at that wavelength that has passed through a selectively absorbing substance) of plasmonic particles is explained. The wavelength of the scattered light is the same as that of the incident light.

Scattering theory was the first approach applied to explain the phenomenon of SPB. Mie applied Maxwell's equations to spherical NPs embedded in a medium and chose adequate boundary conditions to derive the equation for cross-section C_{ext} given by (Mie 1908)

$$C_{ext} = \frac{24\pi^2 R^3 \varepsilon_m^{2/3}}{\lambda} \frac{\varepsilon_2}{(\varepsilon_1 + 2\varepsilon_m)^2 + \varepsilon_2^2} \qquad (2.27)$$

where
   λ is the wavelength of the incident wave
   ε_m is the dielectric constant of the surrounding medium

Scattering cross section is a hypothetical area, which describes the probability or likelihood of light being scattered by a particle. At resonance,

$$\varepsilon_1 = -2\varepsilon_m \qquad (2.28)$$

so that C_{ext} reaches a maximum. For $\varepsilon_m = 1$, Equation 2.28 reduces to

$$\varepsilon_1 = -2 \qquad (2.29)$$

which is the same as Equation 2.23.

2.10   Parameters Controlling the Position of Surface Plasmon Band of Nanoparticles

What are the vital parameters affecting the position of surface plasmon band? Equations 2.16, 2.17, 2.23, and 2.24 suggest that the electron density *n* inside the NP as well as the permittivity of medium surrounding the particle (here ε_0) are critical deciding parameters controlling the dielectric constant of the NP and the resonance frequency in surface plasmon phenomenon, that is, the position of SPB. These properties have been extensively exploited in nanosensor fabrication and therefore require elaborate explanation.

2.10.1 Effect of the Surrounding Dielectric Medium

What are the principal approaches followed to explore this effect? Two approaches are followed here. The first approach is to change the matter surrounding the NP and the second one is to vary the distance separating two NPs. In experiments by Underwood et al. (1994) and Liz-Marzán et al. (1996), it was shown by using gold (Au) NPs stabilized with a polymeric comb (for preventing aggregation of the particles upon solvent transfer) that the SPB shifted toward the lower-frequency and therefore lower-energy red side when the refractive index *n* of the solvent and hence its dielectric constant $\varepsilon_r = n^2$ was increased.

In further experiments (Ung et al. 2001), silica-coated gold NPs were prepared in a core–shell structure. The thickness of the silica (SiO_2) coating was changed to study its effect. In solution, the position of the SPB was found to depend on the thickness of the silica coating and therefore separation between particles.

Then thin films of silica-coated NPs were deposited on glass substrates. The colors of these films were observed to vary with silica thickness, which determined the interparticle distance. As this distance decreased, the color of the film shifted from red to blue. The close distance between the particles causes dipole–dipole interactions (electrostatic interactions of permanent dipoles in molecules by virtue of the existence of partial charges on its atoms, e.g., attractive forces between the positive end of one polar molecule and the negative end of another polar molecule; the interaction energy depends on the strength and relative orientation of the two dipoles, as well as on the distance between the centers and the orientation of the radius vector connecting the centers with respect to the dipole vectors) leading to the SPB shift. This effect is identical to the color change observed on aggregation of NPs in solution. The number of layers of particles in the film was also found to affect the SPB position.

In solutions, chemically induced reversible aggregation and de-aggregation of NPs is possible. Grafted (to attach) gold NPs stabilized with dithiol (D-proline reductase, an enzyme; a compound having two thiol groups, the thiol being a sulfur-containing organic compound having the general formula RSH, where R is another element or radical) could be cleaved or rejoined (Thomas et al. 2004). Upon aggregation of the NPs, a red shift (a proportional decrease in the frequency of light toward lower energy, or longer wavelength) of the solution was noticed.

2.10.2 Influence of Agglomeration-Preventing Ligands and Stabilizers

What is the function of ligands (molecules, ions, or atoms bonded to the central metal atom of a coordination compound; substances, e.g., hormone, drug, functional group, etc., that bind specifically and reversibly to another chemical entity to form a larger complex) *and stabilizers on nanoparticle surfaces?* In order to lower their surface tension, metal NPs in solution have a strong urge to agglomerate thereby forming bigger particles. To ensure NP stability, their surfaces are protected by covering with ligands or stabilizing agents such as amines (e.g., RNH_2 are organic derivatives of ammonia formed by replacing the hydrogen atoms in the ammonia, one at a time, by hydrocarbon groups), thiols (organosulfur compounds structurally similar to alcohols but containing a sulfur atom in place of the oxygen atom normally found in alcohols) or phosphines (PH_3 and compounds derived from it by substituting one, two, or three hydrogen atoms by hydrocarbyl groups R_3P, RPH_2, R_2PH, and R_3P [R≠H] are called primary, secondary, and tertiary phosphines, respectively).

How are the ligands/stabilizers bonded to metal nanoparticles? For bonding purposes, either there is an exchange of electrons between the metal atom of the NP and the ligand through oxidation/reduction, or bonding takes place through electrostatic mechanism. In all cases, the stabilizers form shells surrounding the metal NPs.

How is the SPB position affected? The alteration of electron density n inside the NP resulting from the bonding of the stabilizer with the NP changes the SPB position. It moves toward the lower-frequency red side when n falls.

2.10.3 Effect of Nanoparticle Size and Shape

How does the size of nanoparticle affect the SPB position? From Equation 2.27, it is evident that the cross section of a spherical NP is proportional to its radius. Further, the mean free path of an electron in gold (Au) or silver (Ag) is around 50 nm. Hence, an electron is more likely to encounter the wall of the NP if the NP has dimensions in this range. This disturbs the system dynamics. As a result, the SPB of a noble metal NP is blue-shifted with decreasing particle diameter.

How does nanoparticle shape exercise its influence on SPB? Metal NPs are made in a variety of shapes such as nanospheres, nanorods, nanotriangles, nanooctehedra, etc.; octahedra are polyhedrons with eight faces. These shapes differ in their scattering properties. Generally, reduction of symmetry of the NP shape is accompanied by the appearance of additional modes of resonance, for example, the nanosphere shows one resonance peak while a nanorod has two. Two resonances take place: one for oscillations of electrons along the small axis of the NP (transverse mode) and the second for oscillations along the longer axis (longitudinal); the second one is blue-shifted with respect to the first.

2.10.4 Compositional Effect

Composition of the NP strongly affects the SPB. Cu, Ag, and Au NPs show different behaviors, for example, Ag NPs provide a sharper SPB signal, a different window and four times larger extinction coefficient (a parameter defining how strongly a substance absorbs light at a given wavelength, per mass unit or per molar concentration) (Mulvaney 1996, Cao et al. 2001). A red shift is noticed in the position of SPB of an Ag–Au alloy on moving from pure Ag to pure Au (Zou and Schatz 2004, Zou et al. 2004).

2.11 Quantum Confinement

With what material properties is quantum confinement effect related and in what range of sizes is it observed? Quantum confinement is change of electronic and

optical properties of a material when the material sampled is of sufficiently small size, typically, 10 nm or less.

How does quantum confinement originate? This phenomenon arises from the association of several physical processes or interactions in solids with a characteristic length (a convenient reference length [usually constant] of a given configuration) scale. When one dimension of the solid is comparable to or smaller than this characteristic length, the concerned property becomes sensitive to the size of the solid.

2.11.1 Quantum Confinement in Metals

The reduction of the size and dimensionality of metals results in a drastic change in the electronic properties as the spatial length scale of the electronic motion is reduced with decreasing size. When one dimension of a metallic material becomes comparable to the de Broglie wavelength (the wavelength of a particle, given by $\lambda = h/p$ where h is Planck's constant and p is the momentum), typically fraction of a nm for Cu, the quantum confinement effects appear.

Are quantum confinement effects easily observed in metals? Scattering or incoherent phenomena (in which the waves do not maintain a fixed and predictable phase relationship with each other over a period of time) prevent the observation of these effects in metals in the nano state unless low temperatures or high magnetic fields are used.

2.11.2 Quantum Confinement in Semiconductors

Electrons occupy one of two bands (valence or conduction) in a crystal of a semiconductor material. *Valence band* contains electrons that occupy positions in the crystal lattice, that is, in chemical bonding, and *conduction band* contains those electrons participating in electrical conduction. By providing the proper stimuli, one or more electrons can be encouraged to move from valence band to the conduction band, that is, it is released from the chemical bond and free to move (Figure 2.7). As an electron moves from valence band to the conduction band, it creates a *hole*, which is positively charged. Thus excitation of a semiconductor creates an electron–hole pair:

$$h\nu \Leftrightarrow e^-(\text{CB}) + h^+(\text{VB})$$

(2.30)

where $h\nu$ is the energy of the photon.

Together, the hole and the electron are referred to as a *Wannier–Mott exciton* or simply an *exciton*. The electron and the hole in the exciton normally maintain their distance from each other, usually delocalized over a length much longer than the lattice constant. This distance is called the *exciton Bohr radius*. Delocalization is the spreading or sharing of electron–hole pair
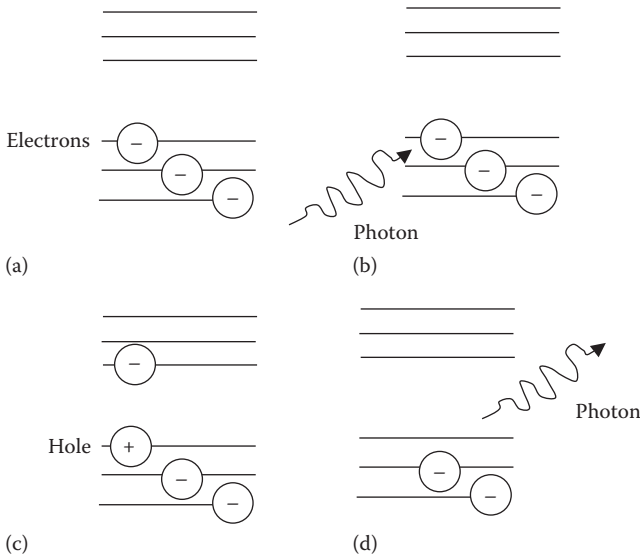
FIGURE 2.7
(a) Energy band diagram of a semiconductor at 0 K showing electrons in the valence band. (b) A photon strikes an electron in the conduction band to create an exciton. (c) Due to the photon strike, an electron is promoted to the conduction band leaving behind a hole in the valence band. An exciton has been produced. (d) The electron drops to the valence band recombining with a hole and emitting light.

between more than two atoms. The exciton is considered an elementary excitation of condensed matter that can transport energy without conveying net electric charge. The electron and hole may have either parallel or antiparallel spins; these spins are coupled by the exchange interaction, producing exciton fine structure.

The hole is thought of as the absence of an electron and acts as a particle with its own effective mass and charge in the solid. The spatial separation of the electron and its hole (in an "exciton") is calculated using a modified Bohr model (a planetary model of an atom consisting of a small, positively charged nucleus orbited by negatively charged electrons similar to the planets orbiting the Sun [except that the orbits are not planar]; the gravitational force of the solar system is mathematically akin to the Coulomb [electrical] force between the positively charged nucleus and the negatively charged electrons) in which the radius of the sphere defined by the three-dimensional separation of the electron–hole pair is given by (Murphy and Coffer 2002)

$$r = \frac{\varepsilon_0 \varepsilon h^2}{\pi m_r q^2} \tag{2.31}$$

where
  ε_0 is the free-space permittivity
  ε is the dielectric constant of the semiconductor
  h is the Planck's constant
  m_r is the reduced mass of the electron–hole pair
  q is the electronic charge

The reduced mass of the electron–hole pair is simply written as

$$m_r = \frac{m_e^* \, m_h^*}{m_e^* + m_h^*} \tag{2.32}$$

where m_e^*, m_h^* are the effective masses of electron and hole, respectively. The effective mass of a particle is the mass it seems to carry in the *semiclassical model* of transport in a crystal. Electrons and holes in a crystal respond to electric and magnetic fields almost as if they were particles with a mass dependence in their direction of travel, an effective mass tensor. Tensors, defined mathematically, are simply arrays of numbers, or functions, which transform according to certain rules under a change of coordinates, extending the notion of scalars, geometric vectors, and matrices to higher orders. Tensors provide a natural and concise mathematical framework for formulating and solving problems in areas of physics.

In a simplified picture that ignores crystal anisotropies, electrons and holes behave as free particles in a vacuum, but with a different mass. This mass is usually stated in units of the ordinary mass of an electron m_e (9.11×10^{-31} kg). In these units, it is usually in the range 0.01–10, but can also be lower or higher. The effective masses of the electron and hole have been determined for many semiconductors by ion cyclotron resonance (a phenomenon related to the movement of ions in a magnetic field; it is used for accelerating ions in a cyclotron, and for measuring the masses of an ionized analyte), and have been found to be in the range of $0.1m_e$–$3m_e$. The radius r has values such as 1 nm for CuCl, 2 nm for ZnS, 3 nm for ZnSe, 4 nm for CdS, 6 nm for CdSe, 8 nm for CdTe or InP, 20 nm for PbS, 34 nm for InAs, 46 nm for PbSe, and 57 nm for InSb (Zimmer 2006).

Example 2.2

Calculate the Bohr radii for CdS ($\varepsilon = 8.4$) and CdSe ($\varepsilon = 9.7$) given that for CdS, the electron and hole effective masses $m_e^* = 0.14m_e$, $m_h^* = 0.51m_e$; and for CdSe, $m_e^* = 0.11m_e$ and $m_h^* = 0.44m_e$ where $m_e =$ electron mass $= 9.1095 \times 10^{-31}$ kg. Also, $\varepsilon_0 = 8.854 \times 10^{-12}$ Fm^{-1}, $h = 6.626 \times 10^{-34}$ J-s, and $q = 1.60219 \times 10^{-19}$ C.
  The Bohr radius is

$$r = \frac{\varepsilon_0 \varepsilon h^2}{\pi \left(m_e^* \, m_h^* / (m_e^* + m_h^*) \right) q^2} \tag{2.33}$$

For CdS,

$$r_{CdS} = \frac{8.854 \times 10^{-12} \times 8.4 \times (6.626 \times 10^{-34})^2}{3.14 \times \left((0.14m_e \times 0.51m_e) / (0.14m_e + 0.51m_e)\right)(1.60219 \times 10^{-19})^2}$$

$$= \frac{3.688 \times 10^{-39}}{9.1095 \times 10^{-31}} = 4.0485 \times 10^{-9}\,\text{m} \qquad (2.34)$$

For CdSe,

$$r_{CdSe} = \frac{8.854 \times 10^{-12} \times 9.7 \times (6.626 \times 10^{-34})^2}{3.14 \times \left((0.11m_e \times 0.44m_e) / (0.11m_e + 0.44m_e)\right)(1.60219 \times 10^{-19})^2}$$

$$= \frac{5.31586 \times 10^{-39}}{9.1095 \times 10^{-31}} = 5.84 \times 10^{-9}\,\text{m} \qquad (2.35)$$

2.11.3 Bandgap Energies

How does the bandgap in a bulk semiconductor differ from that in a small nanosize crystal? In a bulk semiconductor, the bandgap is centered about the atomic energy levels with a width proportional to the nearest-neighbor interactions. The Fermi level is located between the conduction and valence bands and the optical behavior is controlled by the levels near the band edges. The bandgap energies are lower than in smaller-size crystals where quantum confinement effects become perceptible. Bandgap of bulk CdSe is 1.7 eV. On shrinking the crystal size from 6 to 2 nm, the bandgap ranges between 1.9 and 2.8 eV. Likewise, the bandgap of CdS is 2.4 eV. By dimensional shrinkage, it is varied between 2.5 and 4 eV.

What is the explanation for bandgap enlargement as crystal size is shrunk? The increase of bandgap with decreasing crystal size is considered as the energy cost of confining the exciton within dimensions smaller than the Bohr radius. More clear interpretation is provided by the particle-in-a-box model of electron discussed in the ensuing section.

2.11.4 Bandgap Behavior Explanation by Particle-in-a-One-Dimensional Box Model of Electron Behavior

To understand the effect of confining the particles within defined boundaries on bandgap energy, let us recall one of the simplest models for electron behavior in elementary quantum mechanics (a mathematical machine for predicting the behaviors of microscopic particles on the basis of their dual particle-like and wave-like behavior, departing from classical mechanics primarily at the atomic and subatomic scales), that of particle in a one-dimensional box (Figure 2.8). (The particle in a one-dimensional box.)

What are the salient features of this model and the prescribed particle energies? An electron is allowed to move along an arbitrary spatial dimension, the
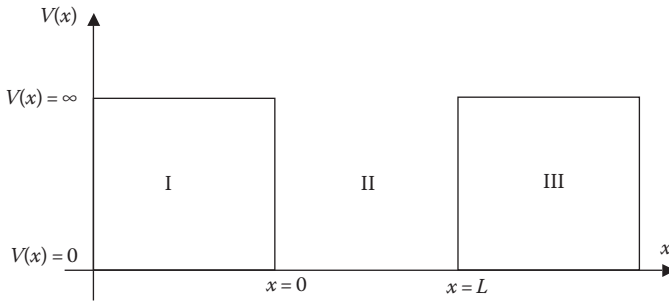
FIGURE 2.8
Particle in a one-dimensional box model (http://user.mc.net/~buckeroo/PODB.html).

x-axis, suppose. The x-axis is divided into three regions, I, II, and III where {I | $x<0$}, {II | $0<x<L$}, and {III | $x>L$}. In regions I and III, an infinite potential energy barrier is present. The potential energy of any particle within these regions becomes infinite, implying that in order to exist there, the total mechanical energy of the particle must be infinite. Since this will never be true, the particle cannot exist within regions I and III. Hence, it is restricted to region II.

What are possible wave function values in different regions? Clearly, the wave function (a variable quantity that mathematically describes the wave characteristics of a particle; the square of the wave function gives the probability or likelihood for finding the particle at a given point in space and time) of the electron must be zero everywhere except for {$0<x<L$}. Remembering that all wave functions are continuous (the functions whose graphs have no breaks, gaps, or holes), the wave function at $x=0$ and at $x=L$ must be zero; otherwise a discontinuity would occur.

What is the goal of this model and how do we seek it? The objective is to determine the set of possible wave functions associated with the electron, and then to use such wave functions to determine the energy eigenvalues (German word *eigen* meaning "characteristic" or "unique"; discrete or quantized energies). One should pay attention to the reason why such discrete (step-like) eigenvalues appear. This is because a particle motion accompanies the wave function. In other words, the eigenstates with discrete energy eigenvalues occur because of the *particle–wave duality*. This is extremely surprising and never happens in the classical theory.

To begin, one looks at the time-independent Schrödinger equation.

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} = E\psi \tag{2.36}$$

neglecting the potential energy term.

Why is the time-dependent wave equation not used? The time-dependent form is not necessary since the states of the electron and its surroundings are constant.

For solving Equation 2.36, it is rearranged into the form

$$\frac{d^2\psi}{dx^2} = -k^2\psi \tag{2.37}$$

by multiplying Equation 2.36 by $(-2m/\hbar^2)$ to obtain

$$\frac{d^2\psi}{dx^2} = -\left(\frac{2mE}{\hbar^2}\right)\psi \tag{2.38}$$

The general trigonometric solution to this second-order differential equation is

$$\psi = N\sin\left(\frac{\sqrt{2mE}}{\hbar}x\right) \tag{2.39}$$

which is the generalized wave function set for the electron.

How are the energy eigenvalues obtained? To determine the energy eigenvalues associated with these wave functions, one looks at the right end point of the one-dimensional box, at $x = L$. Knowing that the wave function is zero at this point, one can write

$$0 = N\sin\left(\frac{\sqrt{2mE}}{\hbar}L\right) \tag{2.40}$$

Since the sine function produces an output of zero when its input values are multiples of π, we can write

$$\frac{\sqrt{2mE}}{\hbar}L = n\pi \tag{2.41}$$

where n is a member of the natural number set $\{1, 2, 3,...\}$.

What is the significance of number n? The number n is called a quantum number, a number which determines the particular state that the electron occupies.

Rearranging Equation 2.41 to isolate energy, we write

$$E = \frac{n^2\pi^2\hbar^2}{2mL^2} = \frac{n^2\pi^2(h/2\pi)^2}{2mL^2} = \frac{n^2h^2}{8mL^2} \tag{2.42}$$

Thus the energy of a particle of mass m ($= m_e$ for electron) confined by one-dimensional box of dimension L of infinite potential is given by the equation

$$E_n = \frac{n^2 h^2}{8 m_e L^2} \tag{2.43}$$

where
   n is the quantum number
   h is Planck's constant

What are the implications of Equation 2.43? Equation 2.43 implies that on decreasing the dimension L of the box, the spacing between the energy levels of the particle in the box increases (Figure 2.9). Thus it can be said that on constraining the charge-carrier particles within reduced dimensions, the quantum confinement effect causes a change in the density of electronic states. For a particle that is confined in a three-dimensional box constrained by walls of infinitely high potential energy, the allowed energy states for the particle are discrete with a nonzero ground state energy. As the length of the box, which corresponds to the radius of the QD, is changed, the energy gap between the ground and the first excited state varies in proportion to $1/r^2$.

For a real nanocrystal, this means that the smaller the particle radius, the larger the energy gap to the first electronically excited state becomes. For a semiconductor material, there exists a size regime bounded by the onset of molecular cluster structure on the smaller side and by the Bohr radius on the larger side in which the bandgap energy varies strongly with the crystal size. As relatively small changes in dimensions produce large changes in bandgap energy, the nanocrystal behaves as a tunable bandgap material, showing variable bandgap controlled by crystal size.
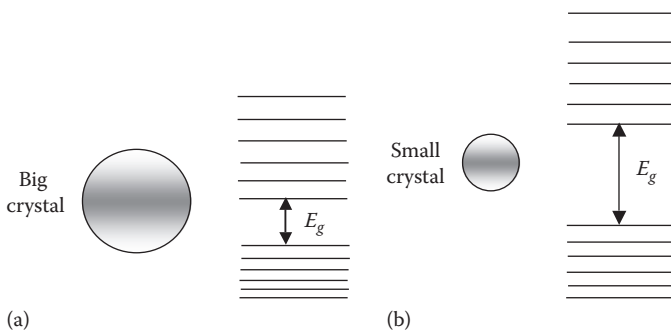
FIGURE 2.9
Effect of dimension of the box on energy-level diagrams of particle in a box: (a) big crystal and (b) small crystal. Note the change in inter-level energy differences as well as bandgap with decrease in size of the semiconductor crystal.

What happens for optically allowed transitions? In case of an optically allowed transition, this relates to a blue shift in absorption and emission as has been observed in all kinds of nanocrystals. This is because the frequency and hence energy of blue light is higher than that for other parts of the spectrum.

Is the tunability of bandgap produced here by altering the chemical composition of the semiconductor? Interestingly, this tunability and tailoring of the bandgap energy and therefore the energy of emitted photons for the fluorescent semiconductors is not brought about through the changes in chemistry of the material, as done in conventional fluorophores; the fluorophores are molecules or parts of molecules that emit fluorescence when excited with light.

What specific advantage is derived from these nanocrystals? A series of fluorophores, with nearly identical reactivities and chemical properties are obtained from the same material through the influence of size changes only. Compositional changes are not required.

Example 2.3

Given that m_e = electron mass = 9.1095×10^{-31} kg, h = Planck's constant = 6.626×10^{-34} J-s, calculate the energy-level differences $(E_2 - E_1)$ for semiconductor nanocrystals of sizes 2 and 20 nm. Do you notice the broadening of the energy-level difference with decrease in size of the nanocrystal?

For $L = 2$ nm,

$$E_1 = \frac{n^2 h^2}{8 m_e L^2 q} eV = \frac{(1)^2 \times (6.626 \times 10^{-34})^2}{8 \times 9.1095 \times 10^{-31} \times (2 \times 10^{-9})^2 \times 1.60219 \times 10^{-19}} = 0.094 \text{ eV}$$

(2.44)

and $E_2 = 0.376$ eV.

Therefore $(E_2 - E_1) = 0.376 - 0.094 = 0.282$ eV. Similarly for $L = 20$ nm,

$$E_1 = \frac{n^2 h^2}{8 m_e L^2 q} eV = \frac{(1)^2 \times (6.626 \times 10^{-34})^2}{8 \times 9.1095 \times 10^{-31} \times (20 \times 10^{-9})^2 \times 1.60219 \times 10^{-19}} = 0.00094 \text{ eV}$$

(2.45)

and $E_2 = 0.00376$ eV. Therefore $(E_2 - E_1) = 0.00376 - 0.00094 = 0.00282$ eV. It is found that $E_2 - E_1$ increases 100-fold for 2 nm crystal as compared to the 20 nm crystal.

Example 2.4

Calculate the energies of the levels $E_1, E_2, E_3, \ldots E_n$ for $L = 1, 2, 5,$ and 10 nm. Hence show the changes in density of electronic states and bandgap energies for electron-in-a-box model of the dimensions specified earlier.

For $L = 1$ nm,

$$E_1 = \frac{n^2 h^2}{8 m_e L^2 q} eV = \frac{(1)^2 \times (6.626 \times 10^{-34})^2}{8 \times 9.1095 \times 10^{-31} \times (1 \times 10^{-9})^2 \times 1.60219 \times 10^{-19}} = 0.376 \text{ eV}$$

(2.46)

Similarly, $E_2, E_3, E_4, E_5, \ldots = 1.504, 3.384, 6.016, 9.4, \ldots$ eV.
For $L = 2$ nm,

$$E_1 = \frac{n^2 h^2}{8 m_e L^2 q} eV = \frac{(1)^2 \times (6.626 \times 10^{-34})^2}{8 \times 9.1095 \times 10^{-31} \times (2 \times 10^{-9})^2 \times 1.60219 \times 10^{-19}} = 0.094 \text{ eV}$$

(2.47)

and $E_2, E_3, E_4, E_5, \ldots = 0.376, 0.846, 1.504, 2.35, \ldots$ eV.
For $L = 5$ nm,

$$E_1 = \frac{n^2 h^2}{8 m_e L^2 q} eV = \frac{(1)^2 \times (6.626 \times 10^{-34})^2}{8 \times 9.1095 \times 10^{-31} \times (5 \times 10^{-9})^2 \times 1.60219 \times 10^{-19}} = 0.01504 \text{ eV}$$

(2.48)

and $E_2, E_3, E_4, E_5, \ldots = 0.06016, 0.13536, 0.24064, 0.376, \ldots$ eV.
For $L = 10$ nm,

$$E_1 = \frac{n^2 h^2}{8 m_e L^2 q} eV = \frac{(1)^2 \times (6.626 \times 10^{-34})^2}{8 \times 9.1095 \times 10^{-31} \times (10 \times 10^{-9})^2 \times 1.60219 \times 10^{-19}} = 0.00376 \text{ eV}$$

(2.49)

and $E_2, E_3, E_4, E_5, \ldots = 0.01504, 0.03384, 0.06016, 0.094, \ldots$ eV.

Tabulating the calculated values $E_1, E_2, E_3, E_4, E_5, \ldots$ for $L = 1, 2, 5,$ and 10 nm, Table 2.4 is obtained. The energy bandgap is modeled by omitting the intermediate energy level (Algar and Krull, 2009). Thus, Table 2.4 becomes Table 2.5. The energy band diagrams corresponding to $L = 1, 2, 5,$ and 10 nm are sketched in Figure 2.10.

TABLE 2.4

Calculated Energy Level Values for Different Box Sizes

Sl. No.	Energy Level Designations	Energy Values in eV for			
		$L = 1$ nm	$L = 2$ nm	$L = 5$ nm	$L = 10$ nm
1	E_1	0.376	0.094	0.01504	0.00376
2	E_2	1.504	0.376	0.06016	0.01504
3	E_3	3.384	0.846	0.13536	0.03384
4	E_4	6.016	1.504	0.24064	0.06016
5	E_5	9.4	2.35	0.376	0.094

TABLE 2.5

Modified Table of Energy Level Values for Different Box Sizes

Sl. No.	Energy Level Designations	Energy Values in eV for			
		$L=1\,nm$	$L=2\,nm$	$L=5\,nm$	$L=10\,nm$
1	E_1	0.376	0.094	0.01504	0.00376
2	E_2	1.504	0.376	0.06016	0.01504
3	E_g	6.106 – 1.504 = 4.602	1.504 – 0.376 = 1.128	0.24064 – 0.06016 = 0.18048	0.06016 – 0.01504 = 0.04512
4	E_4	6.016	1.504	0.24064	0.06016
5	E_5	9.4	2.35	0.376	0.094
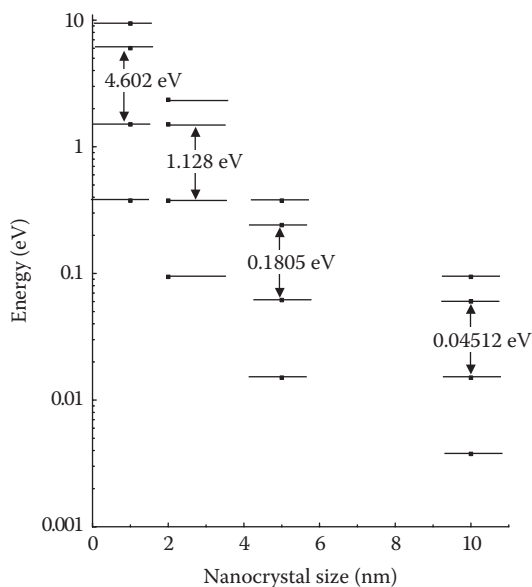
FIGURE 2.10
Energy band diagrams for different L values.

2.12 Quantum Dots

2.12.1 Fundamentals

What are quantum dots and why are they so called? These are colloidal semiconductor nanocrystals, 1–10 nm in diameter, containing typically 10^3–10^4 atoms, so called because quantum confinement takes place here in all three spatial dimensions (Wang and Herron 1991, Bruchez et al. 1998, Murphy and Coffer 2002, Zimmer 2006).

Are quantum dots "artificial atoms"? In some respects but not completely! In a QD, the electrons occupy discrete energy states as they would in an atom. On this argument, the QDs have been called artificial atoms.

Is there a particle-in-a-box picture for quantum dots? Brus (1984) provided the analysis of the particle-in-box picture for QDs based on (i) A Wannier–Mott exciton. (ii) An effective mass approximation for kinetic energy. When an electron is moving inside a solid material, the forces between other atoms affect its motion and it cannot be described by Newton's law. Then the concept of effective mass is introduced to describe the movement of electron in terms of Newton's law. The effective mass can be negative or different depending on the circumstances. Generally, in the absence of an electric or magnetic field, the concept of effective mass is not applicable. (iii) A hydrogenic (hydrogen-like) Hamiltonian. The Hamiltonian H is a mathematical function that can be used to generate the equations of motion of a dynamic system; it equals for many such systems the sum of the kinetic and potential energies of the system expressed in terms of the system's coordinates and momenta treated as independent variables. (iv) Particle-in-a-sphere basis wave functions, resulting in the equation:

Bandgap energy in a quantum dot of radius R - Bandgap energy in bulk

semiconductor = Energy for a particle in a sphere + Coulomb energy between

the electron and hole + Spatial correlation energy between the electron and hole

to reside near the center of the sphere to maximize dielectric stabilization

$$(2.50)$$

Examining Equation 2.33, it is noted that the loosely bound electron and hole have a tendency to reside near each other to maximize their Coulombic interaction. However, the semiconductor dielectric constant is large producing appreciable screening so that the electron and hole tend to reside close to the center of the sphere for dielectric stability. Mathematically, Equation 2.50 is expressed as

$$E_{QD}(R) - E_B = \frac{\hbar^2 \pi^2 \left(m_e^{*-1} + m_h^{*-1} \right)}{2R^2} - 1.786 \frac{q^2}{4\pi\varepsilon_0\varepsilon_r R} - 0.248 E_{Ryd}^* \qquad (2.51)$$

where
  $E_{QD}(R)$ and E_B are the bandgap energies in a QD and bulk semiconductor, respectively
  R is the radius of the QD
  m_e^*, m_h^* are the effective masses of electron and hole
  ε_r is the relative permittivity of the semiconductor material

E^*_{Ryd} is the Rydberg energy of electron–hole pair ($=m_r R_H/\varepsilon_r^2$ where m_r is the reduced mass of electron–hole pair given by Equation 2.32 and R_H is the Rydberg energy of the hydrogen atom = 13.6 eV = the work required to remove an electron from a hydrogen atom; 1 Rydberg [Ry], a subsidiary unit of energy used in atomic physics and optics, is equal to the ionization energy of an atom of hydrogen).

Are there any approximations used in the Brus model? Implicit in this equation is that the QDs are spherical and that the effective masses of charge carriers and the dielectric constant of the solid are constant as a function of QD size.

What is applicability zone of the Brus model? The Brus model maps E_g and size well for larger QDs, but its predictions do not match with experiment well for very small particle size. Although containing the basic physics of the quantum size effect, the model cannot be expected to be quantitatively correct, especially for very small dot sizes, as has been shown by a growing number of experimental studies. This is because for small QDs, the eigenvalues of the lowest excited states are located in a region of the energy band that is no longer parabolic (the breakdown of the effective mass approximation). Many other more complex approximations have been derived theoretically that better match experimentally determined values.

For a quantitative description of the quantum size effect, one should move beyond the effective mass approximation. Better description of the energy band is obtained from the linear combination of atomic orbitals-molecular orbitals (LCAO-MO) (a quantum superposition of atomic orbitals for calculating MO in quantum chemistry) or tight-binding approach (an approach for the calculation of electronic band structure using an approximate set of wave functions, by superposition of wave functions for isolated atoms located at each atomic site) (Wang and Herron 1991). It provides a natural framework to understand the evolution of clusters from molecules to bulk, and the size dependence of the lowest excited-state energy (bandgap).

2.12.2 Tight-Binding Approach to Optical Bandgap (Exciton Energy) versus Quantum Dot Size

First let us briefly review molecular orbital concepts and theory. *Molecular orbitals* are the solutions of the Schrodinger's equation for molecules in the same way as atomic orbitals are solutions of this equation for atoms. Two atomic orbitals overlap to give two MO: the molecular orbital at a lower energy than the overlapping atomic orbitals is called a *bonding molecular orbital* while the molecular orbital at a higher energy than the overlapping orbitals is known as *antibonding molecular orbital*.
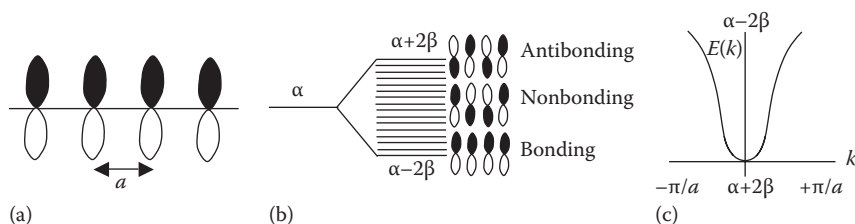
FIGURE 2.11
(a) Orbitals of chain of carbon atoms. (b) Splitting of energy level of orbital into levels for bonding, nonbonding, and antibonding orbitals. (c) Plot of $E(k)$ with k. (After Wang, Y. and Herron, N., *J. Phys. Chem.*, 95, 525, 1991.)

Huckel theory is an approximate molecular orbital theory in which the neighboring atoms are postulated to overlap and bond strongly and equally whereas atoms that are not nearest neighbors are assumed to be noninteracting.

Degenerate level is an energy level of a quantum-mechanical system that corresponds to more than one quantum state. Degenerate states are quantum states of a system having the same energy.

A simple one-dimensional analogue of the bulk solid is provided by an infinite chain of carbon atoms, in which individual atoms are separated by distance a (Wang and Herron 1991), see Figure 2.11a. Moreover, each atom carries one $p\pi$ orbital (polyene, an idealized polyacetylene chain). This infinite chain is equivalent to an N-annulene where N is infinite. Annulene is a class of monocyclic hydrocarbons (having a molecular structure with only one ring) with conjugated double bonds (two or more double bonds separated by single bonds; conjugation is possible by means of alternating single and double bonds). The main ideas behind the quantum size effects of semiconductor QDs are understood by investigating the length dependence of N-annulene or N-polyene; polyene is an unsaturated compound containing more than two double bonds.

What happens in the absence of any interaction between the carbon atoms? In this situation, an N-atom polyene has N degenerate orbitals, ϕ_i, each with on-site energy (Coulomb integral) $\alpha, \langle \phi_i | H | \phi_i \rangle$ where H is the electronic Hamiltonian.

The degeneracy is removed by considering the interaction between ϕ_i and ϕ_j, represented by the transfer integral β, $\alpha, \langle \phi_i | H | \phi_j \rangle$. The lowest energy orbital is situated at $\alpha + 2\beta$ and is bonding between all neighboring atoms (Figure 2.11b); a *bonding orbital* is a molecular orbital formed by a bonding electron whose energy decreases as the nuclei are brought closer together, resulting in a net attraction and chemical bonding. The highest energy orbital is located at $\alpha - 2\beta$ and is antibonding between all neighboring atoms; an *antibonding orbital* is an atomic or molecular orbital whose energy increases as atoms are brought closer together, indicating a net repulsion rather than

a net attraction and chemical bonding. In the middle at the energy α, there is a nonbonding orbital. The energy levels of these orbitals, E_j, are obtained easily on the Huckel level as

$$\text{Annulene}: E(j) = \alpha + 2\beta \cos\left(\frac{2j\pi}{N}\right) \quad j = 0, \pm 1, \ldots,$$

$$(\pm N/2 \text{ for } N \text{ even}) \text{ or } \{\pm (N - 1)/2 \text{ for } N \text{ odd}\}$$

$$(2.52)$$

$$\text{Polyene}: E(j) = \alpha + 2\beta \cos\left(\frac{j\pi}{N+1}\right) \quad j = 1, 2, \ldots, N \qquad (2.53)$$

By defining a new index k, these equations are rewritten in a different form

$$E(k) = \alpha + 2\beta \cos(ka) \qquad (2.54)$$

$k=0$ to ($\pm\pi/a$ for annulene) or (or π/a for polyene). Here $k=2j\pi/Na$ for annulene and $j\pi/(N+1)a$ for polyene; it is called the *wave vector*, a vector which has magnitude equal to the wave number (the number of waves per unit distance in a series of waves of a given wavelength) and points in the direction of propagation of the wave.

By plotting $E(k)$ versus k curve, Figure 2.11c, the dispersion of the energy band is obtained. For an infinite chain, the energy band is continuous from $\alpha + 2\beta$ to $\alpha - 2\beta$ with a width of the energy band $= 4\beta$ but for a chain of finite atoms, the eigenvalues are discrete.

For linear polyene, there is no degeneracy (k varies from 0 to π/a), so it is easily proven by filling electrons into these orbitals, that the HOMO–LUMO (highest occupied molecular orbital–lowest unoccupied molecular orbital) gap increases with diminishing chain length, N.

There is no bandgap in an idealized polyene chain. By introducing two-atom repeat unit and Pierels distortion (Jahn–Teller distortion) into the chain, a bandgap (HOMO–LUMO gap) is created in the middle of the dispersion curve at $k=\pi/a'$ (Figure 2.12). *Jahn–Teller* effect is the distortion of nonlinear molecules or complexes for avoiding degenerate electronic states. For N-annulene, the energy levels are given by

$$E(k) = \alpha \pm \sqrt{\beta_1{}^2 + \beta_2{}^2 + 2\beta_1\beta_2 \cos ka'} \qquad (2.55)$$

where β_1, β_2 are transfer integrals used to describe interactions of orbitals separated by different distances. At $k=\pi/a'$, the bandgap $= 2(\beta_1 - \beta_2)$.

When an electric field is applied, the electron or hole moves as if it has an effective mass of $h^2/(d^2E(k)/dk^2)$. Near the bottom of the band where the $E(k)$ versus k curve is close to parabolic, the effective mass is $m^* = \hbar^2(\beta_1 + \beta_2)/\beta_1\beta_2\, a^2$.
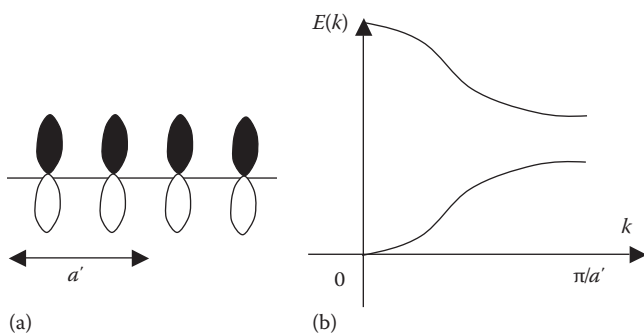
FIGURE 2.12
(a) Definition of *a'*. (b) Bandgap creation. (After Wang, Y. and Herron, N., *J. Phys. Chem.*, 95, 525, 1991.)

For a Peierls-distorted *N*-annulene, the HOMO–LUMO gap also increases with decreasing *N* when *N* is odd, Equation 2.55. But for even *N* the energies of the HOMO and LUMO are independent of *N*. However, on examination of the wave functions, it is noticed that the HOMO and LUMO are two nonbonding orbitals with zero overlap. Hence the actual observable transition takes place at the next LUMO, which is again dependent on *N*.

The LCAO-MO approach presented earlier offers a simple formulation for appreciating important solid-state concepts. The main observation is that with decrease in size of the *N*-annulene or *N*-polyene, its energy levels become discrete and the distance between the eigenvalues, and HOMO–LUMO gaps, increases, as obvious from Equations 2.53 and 2.55. This is essentially the first part of the quantum size effect asserting the increase of the excited-state energy with decreasing QD size.

It may be noted that the curvature of the band and the effective mass of the electron or hole are determined by β. For large values of β, the curvature of the band being steep, the electron or hole effective mass is small, and the width of the band is large. Consequently, a more pronounced quantum size effect is anticipated. In the limit of β approaching the zero limit, as in the cases of most molecular solids (the constituents units are molecules, which do not carry any charge; these molecules are held together by the van der Waals forces), the band is flat and the quantum size effect is absent.

Example 2.5

The forbidden energy gap of CdS (E_B) is 2.4 eV. Find the bandgap energy of a CdS QD of radius $R = 2$ nm, if the electron effective mass $(m_e^*) = 0.14$ electron mass units, hole effective mass $(m_h^*) = 0.51$ electron mass units, free electron mass $(m_e) = 9.11 \times 10^{-31}$ kg, the dielectric constant of CdS (ε_r) is 8.4, electronic charge $(q) = 1.6 \times 10^{-19}$ C, reduced Planck's constant $(\hbar) = 1.054572 \times 10^{-34}$ J-s, Rydberg energy of hydrogen atom (R_H) = 13.6056923 eV.

The exciton Rydberg energy is

$$E_{Ryd} = \left(\frac{m_r}{m_e \varepsilon_r^2}\right) R_H = \left(\frac{m_e/9.104}{m_e \varepsilon_r^2}\right) R_H = \frac{R_H}{9.104 \times \varepsilon_r^2}$$

$$= \frac{13.6056923}{9.104 \times (8.4)^2} = 2.118 \times 10^{-2} \text{ eV} \tag{2.56}$$

where
  R_H is the Rydberg energy of the hydrogen atom $= 13.6056923 \text{ eV}$
  μ is the reduced mass of electron–hole pair given by

$$\mu = \left(\frac{1}{m_e^*} + \frac{1}{m_h^*}\right)^{-1} = \left(\frac{1}{0.14 m_e} + \frac{1}{0.51 m_h}\right)^{-1} = \left(\frac{0.51 + 0.14}{0.14 \times 0.51 m_e}\right)^{-1}$$

$$= \left(\frac{9.104}{m_e}\right)^{-1} = \frac{m_e}{9.104} \tag{2.57}$$

From Equation 2.51,

$$E_{QD}(R) = E_B + \frac{\hbar^2 \pi^2 \left(m_e^{*-1} + m_h^{*-1}\right)}{2R^2} - 1.786 \frac{q^2}{4\pi \varepsilon_0 \varepsilon_r R} - 0.248 E_{Ryd}^*$$

$$= 2.4 + \frac{(1.054572 \times 10^{-34})^2 \times (3.14)^2 \times \left(9.104/(9.11 \times 10^{-31})\right)}{2(2 \times 10^{-9})^2 \times 1.6 \times 10^{-19}}$$

$$- 1.786 \frac{(1.6 \times 10^{-19})^2}{4 \times 3.14 \times 8.854 \times 10^{-12} \times 8.4 \times 2 \times 10^{-9} \times 1.6 \times 10^{-19}}$$

$$- 0.248 \times 2.118 \times 10^{-2}$$

$$= 2.4 + \frac{1.0957857 \times 10^{-36}}{1.28 \times 10^{-36}} - \frac{4.572 \times 10^{-38}}{2.9892 \times 10^{-37}} - 5.2526 \times 10^{-3}$$

$$= 2.4 + 0.8561 - 0.15295 - 5.2526 \times 10^{-3} = 3.098 \text{ eV} \tag{2.58}$$

2.12.3 Comparison of Quantum Dots with Organic Fluorophores

How do quantum dots compete with organic fluorophores? Organic fluorophore-based diagnostics is a versatile and widely practiced technology. QDs provide good substitutes for these fluorophores serving essentially as reporters of information but in a more reliable manner. Advantages as well as disadvantages of organic fluorophores and QDs are listed in Table 2.6.

  The drive to measure more biological indicators simultaneously imposes new demands on the fluorescent probes used in these experiments. To better appreciate the advantages of QDs, let us understand: *What difficulties are*

TABLE 2.6

Organic Fluorophores and Quantum Dots

Sl. No.	Organic Fluorophores	Quantum Dots
1	Narrow and weak absorption spectra	Broad and strong absorption spectra
2	Broad red-tailed photoluminescence (PL) spectra	Narrow and symmetric PL spectra
3	Short fluorescent lifetimes, typically nanoseconds	Longer-lived PL spectra (tens of nanoseconds)
4	Highly susceptible to photobleaching	Resistive to photobleaching
5	Susceptible to chemical degradation	Chemically stable
6	Not optimally suited for multiplexing	Superior spectral properties for multiplexing
7	Smaller in size	Larger in size (hydrodynamic radii ~ 10–15 nm along with surface chemical modification) causing structural or functional perturbation of biomolecules
8	No blinking effect	Luminescence turns on and off upon continuous excitation

faced with conventional fluorophores in a multiplexed experiment? A multiplexed experiment is performed by exciting and observing photoluminescence (PL) from many fluorophores at the same time. Due to the broad red-tailed PL and Stokes shift (the displacement of spectral lines or bands of luminescent radiation toward longer wavelengths than those of the absorption lines or bands; it is the difference in wavelength or frequency units between positions of the band maxima of the absorption and emission spectra) in the range of 10–40 nm of most organic fluorophores, spectral overlapping takes place between fluorophores causing cross talk (a disturbance caused by one communication signal affecting a signal in an adjacent channel) between detection channels. This cross talk is avoided by using spectrally far-spaced dyes. But an organic fluorophore absorbs strongly over only a narrow wavelength range. Hence, when using fluorophores, a separate excitation source is needed for efficient working of each fluorophore. Moreover, rapid photobleaching (photochemical destruction) of the dyes makes their use difficult. Due to the photobleaching effect, the signal drifts over a single measurement or is lost over several measurement cycles. Moreover, conventional dye molecules impose stringent requirements on the optical systems used to make these measurements; their narrow excitation spectrum makes simultaneous excitation difficult in most cases, hindering multicolor real-time imaging (the rapid acquisition and manipulation of information enable images to be produced almost instantaneously). Their broad emission spectrum with a long tail at red wavelengths retards efficient multiplexing (the process where multiple channels are combined for transmission over a common transmission path) by introducing spectral cross talk between different detection channels,

unless accurate emission filters are used to prevent bleed through (often termed crossover or cross talk) to adjacent channels. Finally, synthesis of new molecules is necessary for spectrally shifting the absorption and PL spectra of a fluorophore.

In contrast to the fluorophores, the size of the QD determines its peak PL wavelength. With decreasing size of QD, the peak PL wavelength changes from red to blue. The absorbance onset and emission maximum shift to higher energy. A QD absorbs light at any wavelength smaller than its peak PL wavelength. The excitation tracks the absorbance, resulting in a tunable fluorophore that can be excited efficiently at any wavelength shorter than the emission peak, yet will emit with the same characteristic narrow, symmetric spectrum regardless of the excitation wavelength. This enables the simultaneous excitation of many QDs having different emission wavelengths by a single wavelength source in the blue or ultra violet region, creating a Stoke's shift ~100–300 nm, much larger than for fluorophores. Variation of the material used for the nanocrystal and also of the size of the nanocrystal affords a spectral range of 400 nm–2 μm in the peak emission, with typical emission widths of 20–30 nm (full width at half maximum [FWHM]) in the visible region of the spectrum and large extinction coefficients in the visible and ultraviolet range ($\sim 10^5$ M^{-1} cm^{-1}). The *full width at half maximum (FWHM)* is a parameter commonly used to describe the width of a "bump" on a curve or function; it is given by the distance between points on the curve at which the function reaches half its maximum value. *Extinction coefficient* is a parameter that defines how strongly a substance absorbs light at a given wavelength, per mass unit. Many sizes of nanocrystals may therefore be excited with a single wavelength of light, resulting in many emission colors that may be detected simultaneously. QDs combine a broad excitation spectrum with a narrow emission spectrum.

Do quantum dots have any more advantages over fluorophores? Apart from the aforementioned advantages, QDs offer better or comparable brightness, high photostability (resistance to change under the influence of radiant energy and especially of light) and low reactivity (the rate at which a substance tends to undergo a chemical reaction) than fluorophores. In addition, the fluorescence lifetimes (lifetime refers to the average time the molecule stays in its excited state before emitting a photon) are larger or on the order of tens of nanoseconds, long enough to allow time-gated detection (monitoring the fluorescence of a sample as a function of time after excitation by a flash or pulse of light; short laser pulse acts as a gate for the detection of emitted fluorescence and only fluorescence that arrives at the detector at the same time as the gate pulse is collected, with the detector turned on after excitation and integration of the fluorescence intensity generated) but short enough so that the signal is not limited by low turn over rates (as is the case for lanthanides [14 rare earth chemical elements which lie between lanthanum and lutetium on the periodic table] or metal ligand complexes, the structures consisting of a central atom or ion [usually metallic], bonded

to a surrounding array of molecules or anions [ligands, complexing agents]) with millisecond lifetimes. Fluorescence from single QDs has been observed longer than from most other individual fluorophores enabling to collect a larger number of photons. Further, the synthetic protocol of QDs with different emission wavelengths is simpler and reaction time is the dot size-controlling variable.

What are the disadvantages of quantum dots? Disadvantages of QDs include the toxicity of the component heavy metals Cd or Pb used in II–VI semiconductor dots, the most widely used QDs. Furthermore, the inorganic nature of QDs renders them inherently insoluble in aqueous media. Also, the larger hydrodynamic radii (the effective radius of an ion in a solution measured by assuming that it is a body moving through the solution and resisted by the viscosity of the solution; if the solvent is water, the hydrodynamic radius includes all the water molecules attracted to the ion) of QDs ~10–15 nm are a serious impediment in their application because they cause disturbances in biomolecular reactions.

Blinking or *photoluminescence intermittency effect* in QDs is their other major shortcoming. It is a nearly universal aspect of nanocrystal luminescence, yet it is poorly understood. When it was first observed, "blinking" was attributed to rare photoionization (the physical process in which an incident photon ejects one or more electrons from an atom, ion, or molecule) events, that is, the events in which optically excited electrons tunnel to a nearby trap state.

2.12.4 Types of Quantum Dots Depending on Composition

How are quantum dots classified on the basis of composition? QDs are classified as primary (the first stage), binary (composed of two parts or two pieces), and ternary (composed of three items). *Primary QDs* are those made of a single element such as silicon QDs. *Binary QDs* are those composed of two elements, for example, CdS, CdSe, etc. *Ternary QDs* contain three elements like CdSeTe, ZnCdSe, etc.

2.12.5 Classification of Quantum Dots Based on Structure

How are quantum dots classified from the viewpoint of their structures? The use of semiconductor nanocrystals in a biological context is potentially problematic because the high surface area of the nanocrystal might lead to reduced luminescence efficiency (ratio of energy of emitted photon to that of exciting photon assuming that every exciting photon yields an emitted photon) and photochemical degradation. As already pointed out in reference to NPs in general, bandgap engineering (controlling or altering the bandgap of a material by controlling the composition) concepts borrowed from materials science and electronics have led to the development of core–shell nanocrystal samples with high, room temperature quantum yields (the ratio of the number of photons emitted to the number of photons absorbed ~50%) and
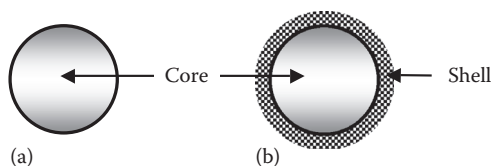
FIGURE 2.13
Structures of QDs: (a) core structure and (b) core–shell structure.

much improved photochemical stability. By enclosing a core nanocrystal of one material with a shell of another having a larger bandgap, one can efficiently confine the excitation to the core, eliminating nonradiative relaxation pathways and preventing photochemical degradation. Thus the shell helps in isolating the carriers to the core, moving the particle surface away from the electron and hole spatially. This separation helps in avoiding the interaction of carriers with surface defects, and trapping of the electron/hole in an intra-bandgap energy level, thus preventing radiative recombination of carriers and consequent lowering of the quantum yield or disallowing the production of broad red-shifted emission from a deep trap state far away from the Gaussian emission of the band edge. Provision of chemical protection to the core, particularly against oxidation, is another advantage of the shell.

Thus there are two classes of QDs (Figure 2.13): *Core nanocrystals* consisting of a single semiconductor material and *core–shell nanocrystals* in which the core semiconductor material is surrounded by another semiconductor of higher energy gap. Generally, the shell material improves the PL quantum yield and stability of the QD but the emission wavelength of the QD is determined by the size of the core semiconductor. Core/shell structures confine the exciton, typically increasing steady-state luminescence (a form of cold body radiation: emission of light by certain materials when they are relatively cool) and decreasing carrier trapping on the nanocrystal surface, and also reducing the blinking rate.

What are type I and type II quantum dots? In *quantum dots of type I*, the shell has generally higher bandgap than the core with the conduction and valence bands of the shell material lying at higher and lower energies, respectively, with respect to vacuum. However, growth of shells having overlapping or interpenetrating conduction or valence band energies is also possible. Such QDs are labeled as type II QDs. *Type II quantum dots* fluoresce at energies lower than the bandgap energies of either core or shell material. Some QDs contain more than one shell, with the first shell an alloy of core and a small/large band offset material and a third outermost shell of a higher bandgap material.

What kind of matching of properties is desired between the core and the shell, and how is it achieved? In order that the shell completely passivates the dangling bonds (unsatisfied chemical bonds associated with an atom in the surface layer of a solid that do not join the atom with a second atom but extend in

the direction of the solid's exterior) on the core surface and there are no gaps separating the islands of shell material causing exposure of core to environmental species, there should be crystal lattice matching between the core and the shell. This is achieved by growing the shell epitaxially on the core, acquiring the crystalline structure and lattice spacing of the core. Therefore the lattice constants of the core and the shell should be nearly same, differing maximum by 5%.

2.12.6 Capping Molecules or Ligands on the Surfaces of Quantum Dots

What are the roles of caps or ligands on QD surfaces? The caps or ligands serve threefold purposes: (i) Partial passivation of dangling bonds on the surfaces of QDs. (ii) Preventing agglomeration of QDs through steric hinderance (the prevention or retardation of inter- or intramolecular interactions by the blockage of access to a reactive site by nearby groups; it arises from the crowding resulting from spatial structure of a molecule). (iii) Imparting solubility to the QDs. Solubility of the QD is not necessarily the same as that of the ligand. The ligands generally have two functional groups, one group binding with the QD and the other group interacting with the environment. A QD coated with surfactant (compounds lowering the surface tension of a liquid) molecules presents hydrocarbon chains to the environment while the polar head groups of the surfactant associate with the QD surface. Consequently, a surfactant-coated QD is water insoluble but is easily dispersed in nonpolar organic solvents such as toluene (C_7H_8 or $C_6H_5CH_3$) or hexane (C_6H_{14}). Polarity refers to a separation of electric charge leading to a molecule or its chemical groups having an electric dipole or multipole moment.

Are there any specific points to be noted for ligands? It must be emphasized that: (i) Ligands allow the tuning of hydrodynamic diameters of QDs, that is, their sizes in water, within wide limits. (ii) The thicker the ligand, the more stable is the QD to degradation. (iii) Ligands and the properties imparted by them are interchangeable after the synthesis of QDs through "ligand-exchange procedures." Simultaneous optimization of all the characteristics is difficult so that compromising optimal solutions must be aimed at, for example, maximizing stability requires a thick ligand shell but a thicker shell may make the QD unusable in applications where size restriction is imposed.

2.13 Carbon Nanotubes

2.13.1 What Are Carbon Nanotubes?

Conceptually, carbon nanotubes are seamless (continuous, flawless) hollow cylinders made by rolling sheets of graphene (an individual graphite layer),

although they are not made in this manner. Figure 2.14 shows a graphene sheet (Ebbesen 1997, Dresselhaus et al. 2001, O'Connell 2006, Dervishi et al. 2009).

How are carbon nanotubes classified? Carbon nanotubes are divided into two main classes: single-wall carbon nanotubes (SWCNTs) (Figure 2.15) and multiwalled carbon nanotubes (MWCNTs) (Figure 2.16). SWCNTs consist of one cylindrical graphene sheet. MWCNTs comprise several nested concentric cylinders with an interlayer spacing of 0.34–0.36 nm. They can be considered as coaxial assemblies of cylinders of SWCNTs.

What are the ranges of lengths and diameters of carbon nanotubes? The lengths of CNTs, either SWCNTs or MWCNTs, are well over 1 µm with diameters ranging from ~1 nm for SWCNTs to ~50 nm for MWCNTs. Because of their high aspect ratio (the ratio of longer dimension to shorter dimension), they are often referred to as *one-dimensional nanomaterials*.

Covalent bond

Carbon atom

(a)   (b)
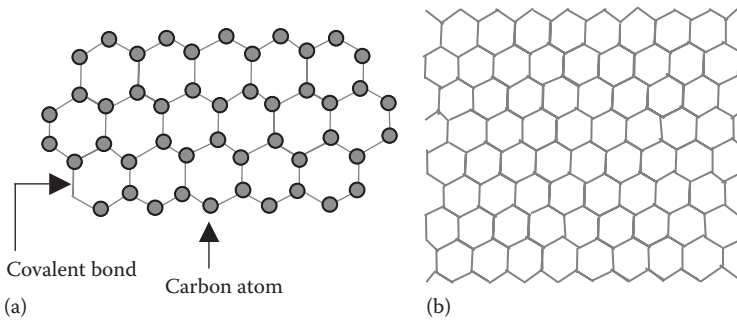
FIGURE 2.14
Atomic arrangement in a graphene sheet: (a) carbon atoms connected by covalent bonds in graphene and (b) simpler atomic representation of graphene.

(a)   (b)
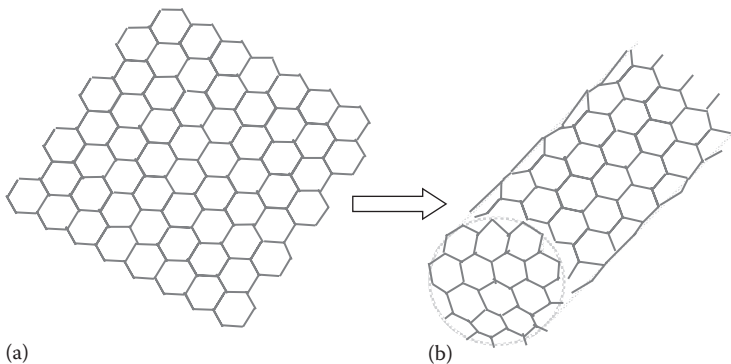
FIGURE 2.15
(a) Graphene and (b) SWCNT.
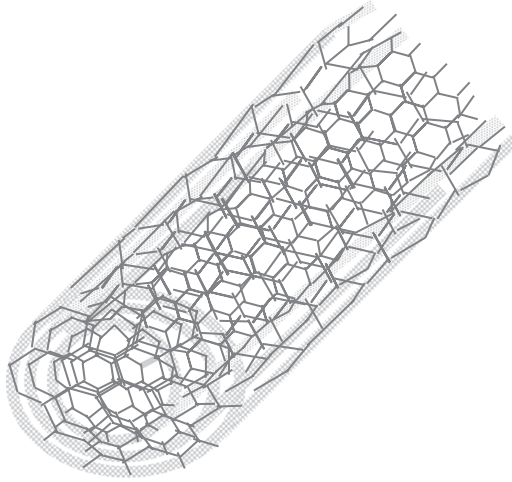
FIGURE 2.16
A MWCNT.

2.13.2 Structure of Graphene

What is the importance of the structure of graphene? Because a CNT is essentially rolled-up graphene, it is interesting to examine the structure of graphene, as shown in Figure 2.15a.

What are the special features of the structure of graphene and how are conducting properties of graphite explained? The honeycomb-shaped graphene is formed by hexagons of carbon atoms bonded together through sp^2 hybrid bonds resulting in interlocking hexagonal carbon rings. Each carbon atom has an extra electron, that is, one electron more than the number of atoms to which it is bonded. The outer-shell electrons of each carbon atom form three in-plane σ bonds and an out-of-plane π-bond (orbital). A σ-bond is the strongest type of covalent chemical bond. In a σ-bond, the electron pair occupies an orbital—a region of space associated with a particular value of the energy of the system—located mainly between the two atoms and symmetrically distributed about the line determined by their nuclei. A π-bond is a covalent bond where two lobes of one involved electron orbital overlap two lobes of the other involved electron orbital. It is a cohesive interaction between two atoms and a pair of electrons that occupy an orbital located in two regions roughly parallel to the line determined by the two atoms. The out-of-plane π-orbital or electron is delocalized and distributed over the entire graphene plane. This freedom of movement of the electron makes graphite thermally and electrically conducting.

Why is graphite a good lubricant and drawing tool? While the carbon atoms in the graphene plane are tightly bonded to each other through covalent linkages, the graphene sheets are feebly held together by the van der Waals

forces. As a result, the graphene sheets slide easily across each other making graphite a good lubricant and drawing tool. For the same reason, graphite of the pencil is not so strong although a layer of graphite is very strong.

2.13.3 Structure of SWCNT

The infinitely many possibilities of rolling a sheet into a cylinder lead to a variety of diameters and microscopic structures of the nanotubes. A diversity of possible configurations is found in practice, and no particular type is preferentially formed.

What is chiral angle? Figure 2.17 shows the unrolled two-dimensional graphene sheet. The translational vector $\vec{T}$, chiral or roll-up vector $\vec{C}_h = m\hat{a}_1 + n\hat{a}_2$, the chiral angle θ, and the unit vectors $\hat{a}_1$ and $\hat{a}_2$ are indicated; m, n are integers. *Chiral angle*, θ, is the angle between the axis of its hexagonal pattern and the axis of the tube, that is, the angle subtended by $\vec{C}_h$ relative to the direction defined by $\hat{a}_1$.

What are the different structures of carbon nanotubes? Armchair, zigzag, and chiral or helical. *How are these structures defined in terms of the tube axis and C–C bond directions?* When the vector $\vec{T}$ (the tube axis) is perpendicular to the C–C bonds, located on opposite sides of each hexagon in the graphene sheet, the structure is referred to as "armchair" because of the ⌄⌄⌄ shape perpendicular to the nanotube axis (Figures 2.18 and 2.19a).

If the tube axis $\vec{T}$ is parallel to the C–C bonds, which are located on opposite sides of the hexagons, the "zigzag" structure is obtained (Figures 2.18 and 2.19b). It is named after the ⋁⋁ shape perpendicular to the axis of the carbon nanotube.
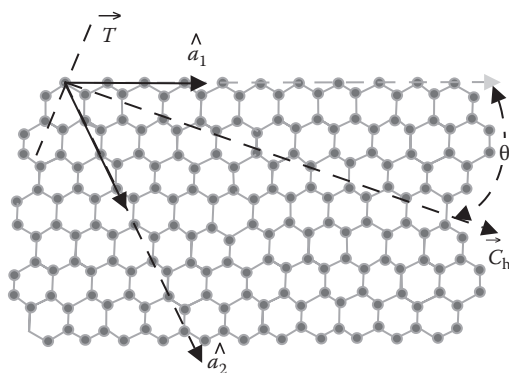
FIGURE 2.17
Chiral vector and its unit vector components. (After Dervishi, E. et al., *Particul. Sci. Technol.*, 27, 107, 2009.)
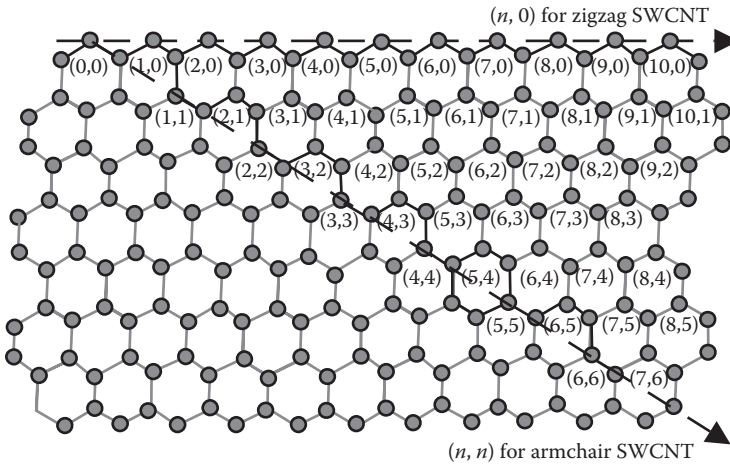
FIGURE 2.18
Chiral vector directions for zigzag and armchair SWCNTs. (After Dervishi, E. et al., *Particul. Sci. Technol.*, 27, 107, 2009.)
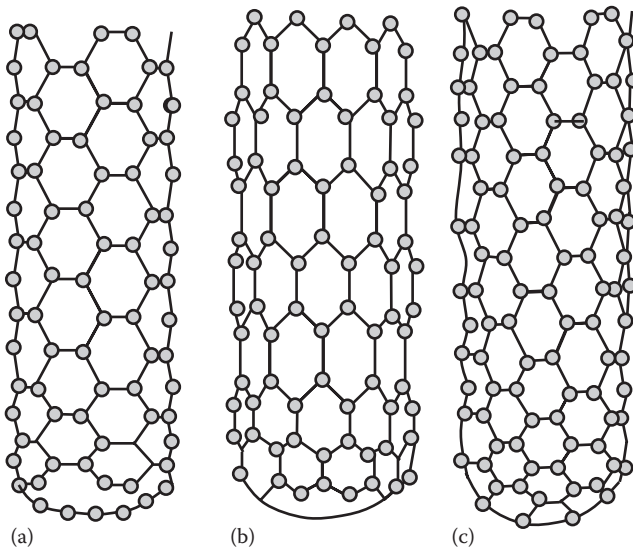
FIGURE 2.19
Carbon nanotubes: (a) armchair, (b) zigzag, and (c) chiral. (After Dervishi, E. et al., *Particul. Sci. Technol.*, 27, 107, 2009.)

It is also possible to roll up the sheet in a direction that differs from a symmetry axis. In the arrangements when the vector $\vec{T}$ lies at an angle with respect to the C–C bonds, structures known as "chiral" or "helical" are formed in which the equivalent atoms of each unit cell are aligned on a spiral (Figure 2.19c).

How are the values of integers m, n related for the three structures? What are the corresponding chiral angle values? The "armchair" CNTs are formed when $m = n$, represented as (n, n) and the chiral angle is 30°, whereas "zigzag" nanotubes are formed when either n or m is zero (which can be written as $(n, 0)$ or $(0, m)$) and the chiral angle is 0°. The rest of the nanotubes $(0 < \theta < 30°)$ are called "chiral" nanotubes, that is, all nanotubes with general (n, m) value and a chiral angle intermediate between 0° and 30° are known as chiral nanotubes.

Since the chiral angle is 0° for zigzag tubes, it is also the angle between the chiral vector and the zigzag direction. It is expressed as (Dervishi et al. 2009):

$$\theta = \tan^{-1}\left\{\frac{\sqrt{3}n}{2m + n}\right\} \qquad (2.59)$$

and the diameter of CNT is given by

$$d = \left(\frac{a}{\pi}\right)\sqrt{m^2 + mn + n^2} \qquad (2.60)$$

where a, the C–C bond length in graphene = 0.246 nm.

Are the ends of SWCNTs open or closed? Pristine (in original, pure state) SWCNTs are generally closed at both ends. A SWCNT has a well-defined spherical tip but the shape of a MWCNT cap is more polyhedral (a geometric solid in three dimensions with flat faces and straight edges) than spherical.

What kind of defects are noticed in CNT structures? Defects are observed in the hexagonal (a polygon with six edges and six vertices) lattice usually occurring in the form of pentagons (any five-sided polygon) and heptagons (or septagon is a polygon with seven sides and seven angles). Pentagons produce a positive curvature of the graphene layer and are frequently found at the cap. Heptagons create a negative curvature of the tube wall.

2.13.4 Mechanical Properties of CNTs

CNTs have properties that distinguish them from other allotropes of carbon; allotropes are different molecular structure modifications of the same element, and can exhibit quite different physical properties and chemical behaviors. *Tensile strength or breaking strain* of individual MWCNTs, the amount of force which a specimen can withstand before tearing, is about 50 GPa (1 GPa = 10^9 Pa; Pascal = newton per m²), which is 20 times stronger than steel (Seetharamappa et al. 2006).

What is the reason for this enormous tensile strength? This strength results from the covalent sp² bonds formed between the carbon atoms, which is stronger than the sp³ bond in diamonds. CNTs are held together by van der Waals forces in a rope-like structure. Another reason why they are strong is

that they are just one large molecule. Unlike other materials, CNTs do not have weak spots such as grain boundaries (the internal interfaces that separate neighboring misoriented grains or single crystals in a polycrystalline solid) in steel.

Young's modulus (Y) of a material is the ratio of longitudinal stress to strain in that direction, a measure of how stiff or flexible an isotropic elastic material is, that is, the material's tendency to deform elastically when a force is applied to it. The value of Y for a material is therefore one of the most important properties in engineering design. Young's modulus of CNTs is independent of tube chirality but depends on its diameter.

How does the Young's modulus of CNTs vary with tube diameter? SWCNTs with diameters between 1 and 2 nm have a very high Young's modulus of about 1 TPa (10^{12} Pa) whereas MWCNTs can have a Young's modulus as high as 1.2 TPa. For comparison, the Young's modulus of steel is only about 0.21 TPa. However, the Young's modulus of SWCNTs decreases from 1 TPa to 100 GPa when the diameter of an SWCNT bundle increases from 3 to 20 nm.

2.13.5 Electrical, Electronic, and Magnetic Properties of CNTs

Depending on their diameter and the chirality, SWCNTs are either metallic or semiconducting. Some nanotubes have conductivities higher than that of copper, while others behave more like silicon.

When are SWCNTs metallic or semiconducting? SWCNTs are metallic when $2n + m = 3i$, where i is an integer; otherwise they are semiconducting.

What is the electrical behavior of armchair and zigzag tubes? All "armchair" carbon nanotubes are metallic because they satisfy the previous equation, while the "chiral" and "zigzag" nanotubes can be either metallic or semiconducting.

What typically is the ratio of metallic to semiconducting SWCNTs? What are representative bandgap values? Approximately one-third of the SWCNTs are metallic and two-thirds are semiconducting. Metallic SWCNTs show a bandgap of 0 eV (a unit of energy = equal to the energy acquired by an electron falling through a potential difference of 1 V = 1.602×10^{-19} J) and are better conductors than metals. The semiconducting SWCNTs have a bandgap of 0.4–0.7 eV.

What is effect of a magnetic field on bandgap? Under the influence of a large magnetic field, the bandgap of semiconducting CNTs can be slightly decreased.

How are CNTs doped? The main methods of doping CNTs are interstitial doping and substitutional doping. In *interstitial doping*, the compound is formed by atoms occupying empty spaces between atoms in a CNT lattice but not at regular lattice sites so that the lattice remains the same. In *substitutional doping*, the dopant atoms replace the carbon atoms and form sp^2 bonding in the CNT structure. Boron and nitrogen are used for *P*- and *N*-type doping respectively. By doping, an order of magnitude increase in the electrical conductivity is observed.

In what direction does the current flow in a CNT? The dominant conduction path in a CNT is along its axis. This is because in the radial direction electrons are confined by the monolayer thickness of the graphene sheet. If the wavelength of the electron is not a multiple of the circumference of the nanotube, it will destructively interfere with itself. Only the wavelengths that are integral multiples of the nanotube circumference will exist.

Why are CNTs called quantum wires? CNTs are often referred to as one-dimensional "quantum wires" (electrically conducting wires in which quantum effects are affecting transport properties) due to the quantum confinement effect on the nanotube circumference.

A quantum wire is a nanostructure, having diameter of the order of a nanometer (10^{-9} m). Alternatively, quantum wires can be defined as structures that have a thickness or diameter constrained to tens of nanometers or less and an unconstrained length. At these scales, quantum mechanical effects prevail, which is the reason for coining the term "quantum wires." Many different types of quantum wires exist, including metallic (e.g., Ni, Pt, Au), semiconducting (e.g., Si, InP, GaN, etc.), and insulating (e.g., SiO_2, TiO_2). An example of organic quantum wires is DNA.

2.14 Inorganic Nanowires

A variety of inorganic materials have been prepared in the form of nanowires with a diameter of a few nm and lengths going up to several microns. In order to produce the nanowires, both vapor-growth and solution-growth processes have been made use of (Meyyappan and Sunkara 2010). Examples of materials used for nanowires in sensor fabrication are silicon (Si), palladium (Pd), gallium nitride (GaN), and metal oxides such as SnO_2, ZnO, CuO, TiO_2, CeO_2, WO_3, In_2O_3 and V_2O_5. The nanowires dispersed in a suspension are deposited on a substrate and contacts are made by thermal or electron beam evaporation. For single-crystal semiconductor nanowires having diameter less than Bohr radius, quantum confinement takes place affecting electron transport, band structure, and optical properties.

2.15 Nanoporous Materials

Porous materials are of scientific and technological importance because of the presence of voids of controllable dimensions at the atomic, molecular, and nanometer scales (Lu and Zhao 2004, Sayari and Jaroniec 2008). The voids endow these materials with properties to discriminate and interact

with molecules and clusters. Interestingly, the notable feature about this class of materials is about the "nothingness" within them, that is, the pore space.

What are nanoporous materials? Nanoporous materials are those having pore sizes less than 100 nm. Such materials are found in abundance in the natural world, both in biological systems and natural minerals. The walls of animal cells are made of nanoporous membranes. Petroleum industry has been long using nanoporous materials called *zeolites* (microporous, aluminosilicate minerals) as catalysts (Sun et al. 2008). *Aerogels* (derived from a gel in which the liquid component of the gel has been replaced with a gas, aerogels are manufactured materials with the lowest bulk density of any known porous solid) are highly porous materials which can have densities as low as four times that of air. For silica aerogels, the distribution of pore sizes has a peak around 5 nm radius. *Activated carbon*, a form of nanoporous carbon is another interesting material. Advancements in capabilities of viewing and manipulating material at the nano level have enabled the directed design of nanoporous materials in place of their opportunistic availability. The main research challenges in nanoporous materials include the fundamental understanding of structure–property relations and tailor-design of nanostructures for specific properties and applications.

2.15.1 Nanoporous Silicon

Nanoporous silicon with a sponge-like structure and very large surface-to-volume ratio (about 10^3 m^2 cm^{-3}) is attractive for sensor applications.

What is the history of nanoporous silicon? Nanoporous Si is not a new material. Porous silicon (PS) layers formed on monocrystalline silicon substrates by electrochemical etching have been known for many years. However, it has been extensively studied in the last 20 years. Its PL behavior was observed by Canham (1990) who reported the discovery of significant visible PL from PS under ultraviolet illumination. Although first PS layer had been observed by Uhlir (1956) and later by Turner (1958) but only Canham's report about room temperature luminescence aroused significant interest in this material. Later, more detailed studies of this film were done by researchers around the world. Considerable progress has been made following the demonstration in 1990 that highly porous material could emit very efficient visible PL at room temperatures. Since that time, all features of the structural, optical, and electronic properties of the material have been subjected to in-depth scrutiny. The majority of research into PS has focused on observations of and explanations for both PL and electroluminescence from this material, and its potential optoelectronic applications. The refractive index of nanoporous silicon can be changed by light. It can emit acoustic waves through thermal stimulation. Further, it produces a stream of electrons (field emission) in absence of vacuum. *Field emission* is the emission of electrons from a solid due to a large electric field at its surface.

What is electrochemical etching technique? Electrochemical etching is the opposite of electroplating. Instead of metal addition, gradual and uniform erosion removes it. A basic electrochemical-etching system consists of an anode (negatively charged piece) and a cathode (positively charged piece) submerged a certain distance from each other in an electrolytic bath. The wafer to be etched forms the anode; a gold cathode is used. The wafer is back-coated with a metallic layer (usually aluminum) to provide a low-resistance electrical contact. The layer of PS is formed by electrochemically etching the crystalline silicon wafer, employing a mixture of hydrofluoric acid and ethanol as an electrolyte.

How does porous silicon compare with microcrystalline silicon material? Typical PL intensity of PS in the visible region (1.5–2.0 eV) is larger by several orders of magnitude as compared to the monocrystalline silicon, which gives luminescence in near infrared range, (0.7–1) to 5 μm, corresponding to the 1.1 eV energy of monocrystalline Si. Silicon, being an indirect semiconductor, is the dominant material of microelectronics. However, it is a poor light emitter and therefore cannot be used in optoelectronics. PS prepared on Si substrate, shows the high external efficiencies of PL (process in which a substance absorbs photons, i.e., electromagnetic radiation, and then reradiates photons) and electroluminescence (emission of light by a material in response to the passage of an electric current) and is suitable for photonic applications.

How is photoluminescence in porous silicon explained? The origin of PL in PS is still controversial. A few models are suggested for explaining the mechanism of PL. According to the model proposed by Canham (1990), radiative recombination of electron–hole pairs occurs within nanometer silicon wires and their energy gaps become larger than that of bulk Si by quantum confinement effect. This model modified by Koch et al. (1993) suggests that electron–hole pairs are photoexcited in nanometer silicon particles and radiatively recombined via Si intrinsic surface states. Another model (Brandt et al. 1992, Cullis et al. 1997) suggests that luminescence from PS was caused by some special luminescence materials, such as SiH_x complexes, polysilanes (a range of polymers having a backbone of continuous silicon atoms), or SiO_2 instead of being an intrinsic property of nanometer Si. A third model believes that excitation of charge carriers occurs in nanometer silicon particles and the photoexcited carriers transfer into the luminescence centers.

How is the structure of nanoporous silicon described? PS is composed of a silicon skeleton permeated by a network of pores. It is possible to define the characteristics of a particular PS layer in a number of ways. The methods of identification include the average pore and silicon branch widths, porosity, pore and branch orientation, and layer thickness.

How is porous silicon used in nanosensors? For sensing applications, the structures are fabricated by forming a PS layer on a silicon substrate and contacting both the PS surface and the rear face of the substrate. Using these structures, a gas sensor based upon the current variation due to the dipole moment of the gas and a humidity sensor based upon the changing current

with humidity have both been demonstrated. Additionally, applications for PS in biosensing have also been demonstrated, using penicillin (a group of antibiotics derived from *Penicillium* fungi) as an example. Coating the large surface area of the PS with a penicillin-sensitive enzyme causes the capacitance voltage curve of the junction to shift with changing concentrations of penicillin.

2.15.2 Nanoporous Alumina

What is the historical background of nanoporous alumina? Nanoporous alumina was first discovered in the 1950s. For the next 20–30 years, limitations in microscopy and other equipment limited the ability of researchers to evaluate their porosity. It was known to be a poor coating process; so, it was abandoned as a process for anodizing aluminum. This porosity has been rediscovered and has sparked significant interest in its properties. Recent work has been performed to produce thin foils of nanopores but these are generally aimed to be filters for both the detection of biological specimens and for use in fuel cells. These pores can be made uniform over large areas. These films are produced from relatively thick aluminum films and can be obtained as commercially available stand-alone filters.

   What are the common porous alumina sensors? Porous alumina sensors are generally available for humidity sensing. These sensors work by adsorbing water molecules on their surface. This adsorption modifies the dielectric constant thereby changing the sensor capacitance. Anodic nanoporous films are produced using common chemicals and electroplating power supplies. Much attention has been devoted toward nanopore ordering and nanopore shape. Highly ordered pore structures require starting with a foil of around 25 μm in thickness and the use of chromic acid. The chromic acid (a corrosive, oxidizing acid, H_2CrO_4) preparations are used to increase the number of pore initiation steps. Highly ordered pores also require sequencing between an etch solution and the pore-forming solution. Anodic nanoporous alumina is made from deposited aluminum, which holds promise for its direct integration into either bipolar or CMOS integrated circuits.

2.15.3 Nano-Grained Thin Films

Thin-film gas sensors (the film thickness is typically less than 100 nm) are of interest because of their relatively small size and low power consumption. A sensor with a thin film of less than a few hundred nanometers has relatively good sensitivity, but usually shows poor stability due to its weak mechanical strength. Consequently, optimization of the micro- or nanostructure of a tin oxide thin film should enhance sensor sensitivity and its stability. Nanocrystalline porous tin oxide thin films obtained by sol–gel technique have a high CO sensitivity combined with a fast and reproducible response.

2.16 Discussion and Conclusions

This chapter provided glimpses of nanomaterials useful for designing and fabricating nanosensors. The knowledge acquired in this chapter about nanosensor constructional materials paves the way toward easier appreciation of nanosensor enabling techniques and unit processes.

Review Exercises

2.1 Define nanomaterials. How are nanomaterials classified on the basis of their composition and origin or occurrence? Name two nanomaterials used in nanosensor fabrication.

2.2 What is an NP? How many dimensions does an NP have? Give one example each of: (i) a one-dimensional nanostructure and (ii) a two-dimensional nanostructure.

2.3 What is a core/shell NP? Explain how the shell improves the properties of the NP. Classify core/shell NPs and give one example of each class.

2.4 Why do the properties of NPs vary with their sizes? Explain.

2.5 Explain the dependence of properties of NPs on their shapes.

2.6 Why do atoms on the surface of a solid possess more energy than those in the bulk? Define surface energy of a solid.

2.7 What is a plasmon? Mention two popular names of surface plasmon phenomenon observed with metallic NPs. Is surface plasmon phenomenon restricted to metallic NPs only?

2.8 What are evanescent waves? Differentiate SPR from LSPR.

2.9 Write and derive the equation for the plasmon resonance frequency of an NP.

2.10 How does the dielectric medium surrounding an NP affect the position of its surface plasmon band? Do ligands and stabilizers have any influence on SPB position of an NP?

2.11 What factors are responsible for non-observation of quantum confinement effects in metals at room temperature?

2.12 In a thought experiment, a semiconductor crystal is continuously subdivided into smaller pieces. Does this subdivision impact the properties of the semiconductor crystal? At what stage any deviation from bulk behavior is noticed?

2.13 What is an exciton? Define Bohr exciton radius. Write the formula for Bohr exciton radius and explain the symbols used.

2.14 What is the effect of shrinking the size of a semiconductor crystal on bandgap energy? Describe the particle-in-a-box model of quantum mechanics and explain the effect of shrinking crystal size on energy level diagrams and bandgap energy of the crystal.

2.15 What is a QD? Why is it called an "artificial atom." Write the Brus equation for the difference in bandgap in a QD and bulk semiconductor. Explain the symbols used and provide the interpretation of the three terms in the equation. Which is the dominant term in energy change?

2.16 Provide an explanation for the advantages offered by QDs in multiplexed experiments in comparison to the organic fluorophores. Mention one disadvantage of QDs.

2.17 Give one example each of single element, binary, and ternary structure QDs.

2.18 For what purposes are capping molecules or ligands applied on the surfaces of QDs?

2.19 What is graphene? What are the two main classes of carbon nanotubes? What are typically the diameters of single and MWCNTs?

2.20 Why is graphite electrically conducting? What is the reason for graphite to act as a good lubricant?

2.21 Define chiral vector and chiral angle. When the chiral angle is 0°, name the CNT formed. What is the CNT for chiral angle of 30°?

2.22 Why is the tensile strength of CNTs much higher than that of steel? Compare the Young's moduli of CNTs with steel.

2.23 State the mathematical equation on satisfying which SWCNT behaves as metallic.

2.24 Name a p-type dopant for CNTs.

2.25 What are nanoporous materials? Give some examples of such materials.

References

Algar, W. R. and U. J. Krull. 2009. Quantum dots for the development of optical biosensors based on fluorescence. In: *Biosensing Using Nanomaterials*. Merkoci, A. (Ed.), John Wiley & Sons, Inc., New Jersey, pp. 199–245.

Brandt, M. S., H. D. Fuchs, M. Stutzmann, J. Weber, and M. Cardona. 1992. The origin of visible luminescence from porous silicon—A new interpretation. *Solid State Communications* 81(4): 307–312.

Bruchez, M., Jr., M. Moronne, P. Gin, S. Weiss, and A. P. Alivisatos. 1998. Semiconductor nanocrystals as fluorescent biological labels. *Science* 281(25): 2013–2016.

Brus, L. E. 1984. Electron–electron and electron–hole interactions in small semiconductor crystallites: The size dependence of the lowest excited electronic state. *Journal of Chemical Physics* 80: 4403–4409.

Canham, L. T. 1990. Silicon quantum wire array fabrication by electrochemical and chemical dissolution of wafers. *Applied Physics Letters* 57(10): 1046–1048.

Cao, Y., R. Jin, and C. A. Mirkin. 2001. DNA-modified core-shell Ag/Au nanoparticles. *Journal of the American Chemical Society* 123(32): 7961–7962.

Cullis, A. G., L. T. Canham, and P. D. J. Calcott. 1997. The structural and luminescence properties of porous silicon. *Journal of Applied Physics* 82: 909 (57 pp.).

Dahlin, A. B., J. O. Tegenfeldt, and F. Hőők. 2006. Improving the instrumental resolution of sensors based on localized surface plasmon resonance. *Analytical Chemistry* 78(13): 4416–4423.

Dervishi, E., Z. Li, Y. Xu, V. Saini, A. R. Biris, D. Lupu, and A. S. Biris. 2009. Carbon nanotubes: Synthesis, properties, and applications. *Particulate Science and Technology* 27(2): 107–125.

Dresselhaus, M. S., G. Dresselhaus, and P. Avouris. 2001. *Carbon Nanotubes: Synthesis, Structure, Properties, and Applications*. Berlin, Germany: Springer-Verlag, 447 pp.

Ebbesen, T. W. 1997. *Carbon Nanotubes: Preparation and Properties*. Boca Raton, FL: CRC Press, Inc., 296 pp.

Eftekhari, A. 2008. *Nanostructured Materials in Electrochemistry*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 463 pp.

Gubin, S. P. 2009 *Magnetic Nanoparticles*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 466 pp.

Hosokawa, M., K. Nogi, M. Naito, and T. Yokoyama (Eds.). 2007. *Nanoparticle Technology Handbook*. Amsterdam, the Netherlands: Elsevier, 644 pp.

Hutter, E. and J. H. Fendler. 2004. Exploitation of localized surface plasmon resonance. *Advanced Materials* 16(19): 1685–1706.

ICX Nomadics: Overview of surface plasmon resonance (http://www.discoversensiq.com/uploads/file/support/spr/Overview_of_SPR.pdf).

Koch, F., V. Petrova-Koch, T. Muschik, A. Nikolov, and V. Gavrilenko. 1993. *Microcrystalline Semiconductors: Materials Science and Devices*: *Materials Research Society Symposium Proceedings*, Pittsburg, PA, vol. 283, 197 pp.

Li, S., M. L. Steigerwald, and L. E. Brus. 2009. Surface states in the photoionization of high-quality CdSe core/shell nanocrystals. *ACS Nano* 3(5): 1267–1273.

Liz-Marzán, L. M., M. Giersig, and P. Mulvaney. 1996. Synthesis of nanosized gold-silica core-shell particles. *Langmuir* 12(18): 4329–4335.

Lu, G. Q. and X. S. Zhao (Eds.). 2004. *Nanoporous Materials: Science and Engineering*. London, U.K.: Imperial College Press, 912 pp.

Meyyappan, M. and M. Sunkara. 2010. *Inorganic Nanowires: Applications, Properties and Characterization*. Boca Raton, FL: CRC Press, Taylor & Francis Group, 433 pp.

Mie, G. 1908. Contributions to the optics of turbid media, particularly of colloidal metal solutions. *Annalen der Physik* 25(3): 377–445.

Moores, A. and P. L. Floch. 2009. The metal nanoparticle plasmon band as a powerful tool for chemo- and biosensing. In: *Biosensing Using Nanomaterials*. Merkoçi, A. (Ed.), John Wiley & Sons, Inc., New Jersey, pp. 137–176.

Mulvaney, P. 1996. Surface plasmon spectroscopy of nanosized metal particles. *Langmuir* 12(3): 788–800.

Murphy, C. J. and J. L. Coffer. 2002. Quantum dots: A primer. *Applied Spectroscopy* 56(1): 16A–27A.

O'Connell, M. J. 2006. *Carbon Nanotubes: Properties and Applications*. Boca Raton, FL: Taylor & Francis, 319 pp.

Rotello, V. 2003. *Nanoparticles: Building Blocks for Nanotechnology*. New York: Springer, 300 pp.

Sayari, A. and M. Jaroniec. 2008. *Nanoporous Materials: Proceedings of the 5th International Symposium,* Vancouver, British Columbia, Canada, May 25–28, 2008. Hackensack, NJ: World Scientific, 738 pp.

Schmid, G. (Ed.). 2004. *Nanoparticles: From Theory to Application*. Weinheim, Germany: Wiley-VCH Verlag GmbH, 444 pp.

Seetharamappa, J., S. Yellappa, and F. D'Souza. 2006. Carbon nanotubes: Next generation of electronic materials. *The Electrochemical Society Interface*, Summer 2006, 15(2), pp. 23–25 and 61.

Sounderya, N. and Y. Zhang. 2008. Use of core/shell structured nanoparticles for biomedical applications. *Recent Patents on Biomedical Engineering* 1: 34–42.

Sun, J., D. Zhang, Z. He, S. Hovmöller, X. Zou, F. Gramm, C. Baerlocher, L. B. McCusker, A. Corma, M. Moliner and M. J. Díaz-Cabañas, Zeolite structure determination using electron crystallography, In Gedeon, A., P. Massiani, and F. Babonneau (Eds.). 2008. Zeolites and related materials, 2 volume set: Trends, targets and challenges, Volume 174 of Studies in Surface Science and Catalysis, Proc. of the 4th Intern. FEZA conf., Paris, France, 2–6 September 2008, ELSEVIER, Amsterdam, pp. 799–804.

The Particle in a one-dimensional box (http://user.mc.net/~buckeroo/PODB.html).

Thomas, K. G., S. Barazzouk, B. I. Ipe, S. T. S. Joseph, and P. V. Kamat. 2004. Uniaxial plasmon coupling through longitudinal self-assembly of gold nanorods. *Journal of Physical Chemistry B* 108(35): 13066–13068.

Turner, R. D. 1958. Electropolishing silicon in hydrofluoric acid solutions. *Journal of the Electrochemical Society* 105(7): 402–408.

Uhlir, A., Jr. 1956. Electrolytic shaping of germanium and silicon. *Bell System Technical Journal* 35: 333–347.

Underwood, S. M., J. R. Taylor, and W. van Megen. 1994. Sterically stabilized colloidal particles as model hard spheres. *Langmuir* 10(10): 3550–3554.

Ung, T., L. M. Liz-Marzán, and P. Mulvaney. 2001. Optical properties of thin films of $Au@SiO_2$ particles. *Journal of Physical Chemistry B* 105(17): 3441–3452.

Wang, Y. and N. Herron. 1991. Nanometer-sized semiconductor clusters: Materials synthesis, quantum size effects, and photophysical properties. *Journal of Physical Chemistry* 95: 525–532.

Yang, P. (Ed.). 2003. *The Chemistry of Nanostructured Materials*. River Edge, NJ: World Scientific, 386 pp.

Zimmer, J. P. 2006. Quantum-dot based biological materials for biomedical imaging. PhD Thesis submitted to Department of Chemistry, Copyright Massachusetts Institute of Technology, MIT, MA.

Zou, S., N. Janel, and G. C. Schatz. 2004. Silver nanoparticle array structures that produce remarkably narrow plasmon line shapes. *Journal of Chemical Physics* 120: 10871 (5 pp.).

Zou, S. and G. C. Schatz. 2004. Narrow plasmonic/photonic extinction and scattering line shapes for one and two dimensional silver nanoparticle arrays. *Journal of Chemical Physics* 121: 12606–12612.

3

Nanosensor Laboratory

3.1 Introduction

A nanosensor laboratory is a place where nanosensors are designed, fabricated, packaged, and characterized. It also includes the facilities for building interfacing signal processing instrumentation. To cater to physical, chemical, and biological nanosensors, it is a multidisciplinary laboratory encompassing physical and life sciences facilities. Figure 3.1 depicts the conception of a nanosensor laboratory incorporating the various sections indicated earlier. The heart of the nanosensor laboratory is the nanotechnology laboratory wherein the nanomaterials are grown/deposited, characterized, and improved.

3.2 Nanotechnology Division

Nanoparticle synthesis and the study of their size and properties are of fundamental importance.

3.2.1 Synthesis of Metal Nanoparticles

What are the synthetic routes for metal nanoparticles? Nanoparticles are prepared by various synthetic routes, for example, (i) the reduction of metal salts in the presence of suitable low molecular weight or polymeric stabilizers, (ii) electrochemical preparation, (iii) decomposition of organometallic (compounds that contain bonds between carbon and a metal) precursors (compounds that participate in a chemical reaction producing another compound), and (iv) vapor deposition methods. *What is the simplest approach of synthesis?* The simplest and the most commonly used bulk-solution synthetic method for metal nanoparticles involves the chemical reduction of the corresponding metal salts.

3.2.1.1 Gold Nanoparticles

Gold nanoparticles are synthesized by the reduction of tetrachloroauric acid ($HAuCl_4 \cdot nH_2O$, $n = 0$, 3, or 4) with trisodium citrate ($Na_3C_6H_5O_7 \cdot nH_2O$, $n = 0$
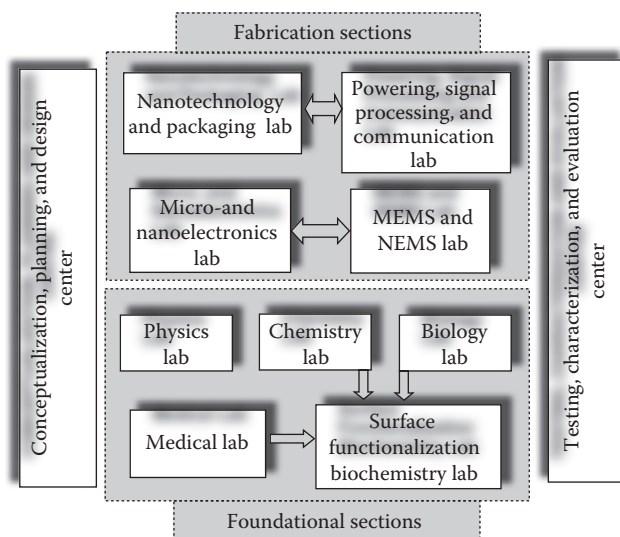
FIGURE 3.1
Divisions of a nanosensor laboratory.

or 2) (McFarland et al. 2004, Ambrosi et al. 2010). The reduction half reaction of gold is given as follows:

$$Au^{3+}(aq) + 3e^- \rightarrow Au(s) \tag{3.1}$$

Materials: Hydrogen tetrachloroaurate (III) trihydrate (HAuCl$_4 \cdot$3H$_2$O) and trisodium citrate (Na$_3$C$_6$H$_5$O$_7 \cdot$2H$_2$O) (s)

Method: (i) All glassware and magnetic stir bars used in the synthesis are scrupulously cleaned in aqua regia (HCl/HNO$_3$ 3:1, v/v) and rinsed in distilled water. (ii) They are oven-dried before use. This precaution is imperative to avoid unnecessary nucleation during the synthesis as well as aggregation of gold colloid solutions. (iii) A 0.25 mM HAuCl$_4 \cdot$3H$_2$O solution (125 mL) is boiled with forceful stirring in a 250 mL round-bottom flask equipped with a condenser; the condenser is an apparatus used to condense vapors to maintain a constant volume of the reaction mixture. (iv) A 40 mM trisodium citrate solution (12.5 mL) is then added quickly to the boiling solution. As the citrate reduces the gold (III), the gold colloid forms slowly, showing a color change from pale yellow to dark red that signifies the formation of Au NPs. (v) The solution is maintained at boiling temperature for 10 min, and then withdrawn from the heating mantle; the heating mantle is a laboratory equipment to apply heat to glass containers without shattering. Stirring is continued for further 15 min. (vi) For detecting the presence of a colloidal suspension, reflection of a laser beam from the particles is observed.

(vii) A small quantity of the gold nanoparticle solution is poured in two test tubes. One tube is used as a color reference and 5–10 drops of NaCl solution are added to the other tube. *Does the color of the solution change as the addition of sodium chloride brings the nanoparticles near each other?* Remember that the formation of nanoparticles earlier was also indicated by a color change. As the particles de-aggregate and move farther apart, the solution turns red. Gold nanoparticle aggregation is accompanied by change in color from red to blue. Red is the color for smaller nanoparticles and blue for larger particles.

3.2.1.2 Silver Nanoparticles

Silver nanoparticles are prepared by the chemical reduction method (Mulfinger et al. 2007) using reduction of silver nitrate by sodium borohydride. The method produces $1.2 \pm 2 \, nm$ particles with plasmon absorbance near 400 nm.

Materials: Sodium borohydride ($NaBH_4$) and silver nitrate ($AgNO_3$).

Method: (i) Thirty milliliters of 0.002 M sodium borohydride ($NaBH_4$) is taken in a flask. A magnetic stir bar is added. The flask is seated in an ice bath on a stir plate. The liquid is stirred and cooled for about 20 min. (ii) Two milliliters of 0.001 M silver nitrate ($AgNO_3$) is poured into the stirring $NaBH_4$ solution at the rate of one droplet per second. No sooner than all the $AgNO_3$ has been added, the stirring is stopped. (iii) As previously mentioned for gold nanoparticles, reflection of a laser beam from the particles aids in perceiving the presence of a colloidal suspension. (iv) A small amount of the solution is transferred to a test tube. When a few drops of 1.5 M sodium chloride (NaCl) solution are added, the suspension turns darker yellow as nanoparticles aggregate. Then it turns gray with increasing aggregation of nanoparticles. *What does this color change suggest?* Color changes are correlated with the aggregation and separation of nanoparticles.

3.2.1.3 Platinum Nanoparticles

The method adopted for preparation of platinum nanoparticles is the reduction of hexachloroplatinate by refluxing a mixed solution of potassium bitartrate ($KC_4H_5O_6$), stabilizer, and hexachloroplatinate. Here, poly(N-vinyl-2-pyrrolidone) [$(C_6H_9NO)_n$, (PVP)], polyethylene glycol [$C_{2n+2}H_{4n+6}O_{n+2}$, (PEG)], and 3,3′-thiodipropionic acid [$C_6H_{10}O_4S$, (TDPC)] are used as the protective agents (Tan et al. 2003).

Materials: Hydrogen hexachloroplatinate (IV) hydrate (H_2PtCl_6, xH_2O), PVP, TDPC, potassium bitartrate, and PEG.

Method: (i) A 20 mL aqueous solution of potassium bitartrate (0.5 wt%) is brought to reflux (~100°C) while stirring, and 20 mL of an aqueous solution

of H_2PtCl_6 (1.0 mM) containing the designated amount of PVP, PEG, or TDPC is added quickly. Reflux means to boil a liquid in a vessel attached to a condenser so that the vapors continuously condense for reboiling. (ii) The concentrations of PVP, PEG, and TDPC are varied (0.5, 1.0, 2.0, and 4.0 wt% (PVP, PEG) and 0.5, 1.0, 2.0, and 4.0 mM (TDPC). This variation is aimed at controlling the size as well as size distribution of Pt nanoparticles. (iii) Brown-colored solutions are obtained after reflux of 20 min and 3–24 h in the presence of PEG and PVP, respectively. The color of the ensuing Pt colloid solutions changes from brown to light brown with increasing concentration ratio of TDPC to H_2PtCl_6.

3.2.1.4 Palladium Nanoparticles

Materials: Potassium tetrachloropalladate (II) (K_2PdCl_4), PVP, TDPC, potassium bitartrate, and potassium hydroxide (KOH) (Tan et al. 2003).

Method: (i) KOH (0.028 g) is dissolved in 20 mL of an aqueous solution of K_2PdCl_4 (1 mM) containing PVP (0.5 wt%) or TDPC (1 mM). (ii) The solution is added to a 20 mL aqueous potassium bitartrate (0.5 wt%) solution whose temperature is raised to ~100°C under constant stirring. (iii) Continued refluxing under stirring for 2 h results in a dark brown–colored solution after the formation of palladium colloid.

3.2.2 Synthesis of Semiconductor Nanoparticles

Silicon nanoparticles (Si NPs) have attracted curiosity because of their unrivaled optical properties. Si NPs are biocompatible. They exhibit high brightness and size-dependent tunable light emission. They also show large photoluminescence quantum efficiency (the average number of electron–hole pairs photoelectrically emitted from a photoreactive surface per incident photon of a given wavelength) and stability against photobleaching (or fading is the photochemical destruction of a fluorophore occurring when a fluorophore permanently loses the ability to fluoresce due to photon-induced chemical damage and covalent modification; it is then considered a photo-degradation product) compared to organic dye molecules. The highly photoluminescence stable and non-aggregated Si nanoparticles provide a competitive, viable replacement for Cd-based quantum dots for biomedical applications. These nanoparticles are prospective non-heavy-element-containing quantum dots well suited for applications in biology. They are useful for electronic and bioimaging purposes.

Rosso-Vasic et al. (2009) synthesized stable and bright emitting amine (organic compounds of nitrogen, such as ethylamine, $C_2H_5NH_2$, that may be considered ammonia derivatives in which one or more hydrogen atoms have been replaced by a hydrocarbon radical)-terminated Si NPs with different alkyl chain lengths between the Si core and amine end group. Their

procedure was as follows: (i) 1.5 g of tetraoctylammoniumbromide, TOAB or TOABr, $[CH_3(CH_2)_7]_4NBr$, molecular formula: $C_{32}H_{68}BrN$, was mixed with 100 mL of dry toluene (C_7H_8 or $C_6H_5CH_3$). The mixture was sonicated (using ultrasound energy to agitate particles in a sample) for 30 min; this sonication was done under a flow of dry argon. (ii) 100 mL of $Si(OCH_3)_4$ (Tetramethyl orthosilicate) was added through a gas-tight syringe and sonication was continued for 30 min allowing entry into the "micelles" (a submicroscopic aggregation of molecules, as a droplet in a colloidal system). (iii) 2.3 mL of $LiAlH_4$ (1 M in THF) was added to the above for the formation of hydrogen terminated Si NPs. (iv) After 30 min sonication, dry and deaerated methanol (30 mL) was poured to react with the surplus $LiAlH_4$; deaeration means removal of air or gas from another gas or liquid. (v) Alkylamine (a compound consisting of an alkyl group attached to the nitrogen of an amine; an example is ethylamine, $C_2H_5NH_2$)-terminated NPs with three different alkyl chain lengths resulted in the reactions of degassed (gas removed) allylamine (2.7 g), hex-5-en-1-amine (2.4 g) and undec-10-en-1-amine ($C1_{11}H_{23}N$) (4.4 g) to each flask with hydrogen-terminated Si NPs obtained under argon (Ar), in the presence of 40 µL of 0.05 M H $2PtCl_6$ catalyst. (vi) After 30 min sonication, 3-aminopropyl, 6-aminohexyl and 11-aminododecyl Si NPs were extracted with water; 3-aminopropyl = $NH_2CH_2CH_2CH_2$; 6-aminohexyl = $C_7H_{16}N_2O_2$, carbamic acid; dodecyl = $CH_3(CH_2)_{10}CH_2$. They were washed with ethyl acetate (EtOAc or EA, $CH_3COOC_2H_5$ or $C_4H_8O_2$) and filtered two times thoroughly by passing through syringe membrane filters. (vii) The resulting Si NPs were further purified by dialysis (the separation of smaller molecules from larger molecules or of dissolved substances from colloidal particles in a solution by selective diffusion through a semipermeable membrane) against water to remove any lingering unreacted aminoalkene (amino = the radical $-NH_2$; alkene = open chain hydrocarbons with one or more carbon–carbon double bonds, having the general formula C_nH_{2n}) and surfactant (a material that can greatly reduce the surface tension of a liquid when used in very low concentrations) that had persisted to remain with the NPs.

3.2.3 Synthesis of Semiconductor Nanocrystals: Quantum Dots

There are two general approaches for the preparation of QDs (Drbohlavova et al. 2009): (i) *Formation of nanosized semiconductor particles through colloidal chemistry*: The colloidal QD synthesis relies on rapid injection of semiconductor precursors into hot and energetically stirred specific organic solvents containing molecules that can coordinate with the surface of the precipitated QD particles. This synthetic route is unsophisticated and can be performed in "one-pot." It constitutes the traditional route for synthesizing quantum dots. (ii) *Epitaxial growth and/or nanoscale patterning*: employing lithography-based technology, this method is widely used in optoelectronics devices such as lasers, infrared photodetectors, etc. However, it will be promising to

use epitaxially grown QDs for *in situ* biosensing, mainly due to the ease of detection.

QD synthesis can be tailored to meet specific requirements, with core, shell, and coating characteristics all affecting photochemical properties. QDs may be manufactured with diameters ranging between broad limits starting from a few nanometers to as large as a few micrometers. Size distributions are controllable within 2% using accurate growth techniques, involving high annealing temperatures (Jamieson et al. 2007).

3.2.3.1 CdSe/ZnS Core/Shell QDs

CdSe–ZnS core/shell QDs are prepared using synthetic techniques involving growth and annealing of organometallic precursors at high temperature. The method according to Wang et al. (2009) is as follows: (i) A mixture of tri-*n*-octylphosphine oxide [TOPO: $OP(C_8H_{17})_3$], a tertiary alkylphosphine, and hexadecyl amine (HDA) ($C_{16}H_{35}N$) is heated. Cadmium acetate $Cd(Ac)_2$ is added to the solution. (ii) The stock solution of trioctylphosphine-selenid (TOP-Se) is prepared by dissolving 0.2 g selenium in 5 g of tri-*n*-octylphosphine (TOP). (iii) This stock solution is injected quickly into the reaction solution along with vigorous stirring of the same. Then nucleation of CdSe nanocrystals takes place. (iv) $Zn(Ac)_2$ and bis(trimethylsilyl) sulfide [$(TMS)_2S$] are added for the inorganic epitaxial growth of the shell on the surface of the core. This epitaxial layer growth is done for about 2 h. (v) CdSe/ZnS QDs with diameter of ~3 nm, thus synthesized, are dissolved in chloroform. They are preserved in a sealed condition.

Riegler et al. (2008) described the following method: cadmium stearate Cd $(C_{17}H_{35}COO)_2$ and TOP-Se are reacted at temperatures above 200°C by fast injection of TOP-Se into a mixture of trioctylphosphine-oxide (TOPO) and cadmium stearate. The CdSe-cores are passivated and annealed by growing a shell of two additional monolayers of ZnS on their surface. Diethylzinc [$(C_2H_5)_2Zn$] and hexamethyldisilathian ($C_6H_{18}SSi_2$) are reacted for 12 h with the CdSe-cores at 160°C in the presence of TOPO and TOP again. The core/shell particles obtained are repeatedly washed with methanol (CH_3OH). They are dispersed in chloroform ($CHCl_3$), and finally stored in 50 mL chloroform as stock solution.

3.2.3.2 CdSe/CdS Core/Shell QDs

In the method reported by Zhai et al. (2011), $Cd(Ac)_2$ and Se powder are selected as source materials. Triethylene glycol ([TREG: triglycol is a colorless odorless viscous liquid, molecular formula = $HOCH_2CH_2OCH_2CH_2OCH_2CH_2OH$] or $C_6H_{14}O_4$) is used as the solvent due to its good hydrophilic feature and high boiling point (288°C). The preparation of CdSe/CdS QDs is done in accordance with the following process steps: (i) 1 g poly(acrylic acid) (PAA, MW = 1800) [$(C_3H_4O_2)_n$] and 0.5 mmol cadmium acetate [$Cd(Ac)_2$] are dissolved into 20 mL

TREG. The solution is heated to 200°C under argon flow. (ii) After 30 min, the solution is cooled down to room temperature, and 19 mg of Se powder is added to it. (iii) The mixture is heated to 240°C. It is kept for a certain period of time such as 1, 5, 60, and 120 min. (iv) As a sulfur source, 19 mg thiourea (CH_4N_2S) is added into the aforementioned CdSe precursor solution. The redundant $Cd(Ac)_2$ in the CdSe precursor solution is used as the source of cadmium. (v) The mixture is heated to 160°C in 1 h. (vi) After reaction for 2 h, the solution is quickly cooled to room temperature. It is precipitated by ethyl acetate. (vii) The resultant solid products are further purified by dialysis and ultrafiltration. Ultrafiltration is a filtration process through a semipermeable medium in which particles of colloidal size are retained by a filter medium while solvent and accompanying low-molecular-weight solutes are allowed to pass through.

3.2.3.3 PbS and PbS/CdS Core/Shell QDs

Silva et al. (2006) used the fusion method to produce PbS quantum dots embedded in S-doped glass matrix $(SiO_2–Na_2CO_3–Al_2O_3–PbO_2–B_2O_3:S)$. The sulfur-doped oxide glass matrix was prepared from high-purity powders using SiO_2 as glass former (an oxide that can readily form a glass) and Na_2CO_3 to decrease the melting point. The mixture was melted in an alumina crucible at 1200°C for 30 min, cooling down to room temperature later. Supplementary thermal treatment of the glass matrix was performed at 500°C to enhance the diffusion of Pb^{2+} and S^{2-} ions. As a consequence of this thermal treatment, PbS quantum dots were formed in the glass matrix.

Pietryga et al. (2008) described a partial cation-exchange method to synthesize PbS/CdSe core/shell QDs in a controlled manner consisting of quenching of the reaction, separation of QDs, and their redispersal for further reaction. For PbS cores with a 5 nm diameter, the process is as follows: (i) A 0.27 g (1.2 mmol) amount of PbO, 0.8 mL (2.5 mmol) of oleic acid $(C_{17}H_{33}COOH)$, 5 mL of TOP (trioctylphosohine) $(C_{24}H_{51}P)$, and 10 mL of 1-octadecene $[CH_2=CH(CH_2)_{15}CH_3]$ were heated to 150°C with continuous stirring under argon flow for 1 h. (ii) The mixture was removed from heat. A solution of 0.1 g (0.6 mmol) of hexamethyldisilathiane $[(CH_3)_3SiSSi(CH_3)_3]$ in 6 mL of 1-octadecene was straight away injected into it. (iii) The reaction was quenched (extinguished) after 10 s by adding 15 mL of cold hexane. (iv) PbS QDs were collected by precipitating with 10 mL of methanol (CH_3OH) and 15 mL of acetone $[(CH_3)_2CO]$, followed by centrifugation (a separation technique that exploits the inherent varied sedimenting property of substances for their isolation by rotating at a very high speed) and removal of the decantate. (v) After redispersing in ~10 mL of toluene, the QDs were washed anew by the same process. They were redispersed once again in 12 mL of toluene. (vi) The suspended cores were subjected to further reaction without initial isolation as a solid because this synthetic method gives essentially quantitative yield.

For CdS shell formation, the procedure is: (i) A 1.0 g (7.8 mmol) amount of CdO, 6 mL (19 mmol) of oleic acid [$CH_3(CH_2)_7CH=CH(CH_2)_7COOH$ or $C_{18}H_{34}O_2$], and 16 mL of phenyl ether ($C_{12}H_{10}O$) were heated to 255°C under argon until all of the CdO had dissolved. (ii) The clear solution was cooled to ~155°C under Ar flow to get rid of water. (iii) The PbS core solution from above was degassed by Ar flow for ~30 min. (iv) Immediately after the temperature of the PbS solution was set to 100°C, the cadmium oleate solution was added via cannula (a flexible tube) accompanied by stirring. (v) Aliquots (portions of solution consisting of equal quantities) were removed and cooled by mixing with hexane. QDs were isolated. They were collected in a fashion analogous to that used for PbS cores.

3.2.4 Synthesis of Metal Oxide Nanoparticles

Many routes have been pursued for SnO_2 nanoparticle growth. These techniques include (i) *thermolysis* (decomposition of a material by heating) of organometallic precursors, (ii) *sol–gel process* (chemical solution deposition: a process in which the aqueous sol (colloidal dispersion) is converted into gel spheres by partial dehydration), (iii) *oxidation* of $SnCl_2$ for tin oxide nanoparticles, (iv) *sonochemistry* (application of ultrasound to chemical reactions), and (v) *hydrothermal synthesis* (the chemical reaction of materials in aqueous solution heated in a sealed vessel).

Nayral et al. (1999) reported the synthesis of tin–tin oxide nanoparticles of low size dispersion (the degree of scatter of data, usually about an average value, such as the median) through a mechanism combining the decomposition of an organometallic precursor (homoleptic tin (II) amides; metal compounds with all ligands identical) and controlled surface hydrolysis (decomposition of a chemical compound by reaction with water) as well as their oxidation into tin oxide nanoparticles without coalescence (uniting into a whole) or change in size.

In numerous applications, the SnO_2 particles appear to grow during sensing operation particularly at high temperature. Baik et al. (2000) studied the stabilization of tin oxide (SnO_2) particles by subjecting a precursor sol solution of SnO_2 to a hydrothermal treatment (hot water). They examined the effects of treating tin oxide gel hydrothermally in an ammonia solution at 200°C for 3 h. Hydrated tin oxide (gel) was precipitated by mixing aqueous solutions of NH_4HCO_3 and $SnCl_4$. After washing with deionized water, a definite amount of precipitated gel (stannic acid, $SnO_2 \cdot nH_2O$), was suspended in an aqueous ammonia solution (NH_4OH) (stannic acid content: 10–40 mass%; stannic acid is acidic form of tin dioxide: any of the series of acids usually occurring as amorphous powders and varying in composition from H_2SnO_3 [alpha-stannic acid] to H_4SnO_4) to undergo a hydrothermal treatment in an autoclave (a strong, pressurized, steam-heated vessel) at 200°C for 3 h. Tin oxide contents of transparent sol solution were calculated as 1.8, 3.2, 6.1, and 8.6 wt%, respectively, from the weight of tin oxide after drying the assured

amount of sol solution (stannic acid contents: 10, 20, 30, and 40 mass%) at 120°C. *What were the sizes of tin oxide particles?* The transparent solutions acquired were used for determining tin oxide particle size with a laser particle distribution analyzer and TEM. The important observation was that the mean crystallite size (5–7 nm) was almost unchanged. On the other hand, the particle size was 4–6 nm in 1.8 wt% sol solution and further, the size of sol particles increased with rising concentration of the consequential sol solution, that is, 5, 8, 10, and 32 nm for 1.8, 3.2, 6.1, and 8.6 wt% tin oxide sol.

Pinna et al. (2004) performed nonaqueous synthesis of nanocrystalline semiconducting metal oxides (tin and indium oxides): (i) Tin (IV) *tert*-butoxide [$C_{16}H_{36}O_4Sn$] (500 mg, 1.216 mmol) or indium (III) isopropoxide [$In_5O(OC_3H_7)_{13}$] (200 mg, 0.685 mmol) was added to benzyl alcohol (abbreviation: BnOH, formula: $C_6H_5CH_2OH$) (20 mL). (ii) The reaction mixture was transferred into a teflon (PTFE: polytetrafluoroethylene $(C_2F_4)_n$) cup of 45 mL inner volume, placed in a steel autoclave, and cautiously sealed. (iii) The autoclave was withdrawn from the glovebox (an enclosed workspace equipped with openings to which are attached gloves for handling, especially dangerous materials in the interior). It was heated in a furnace at 220°C for 2 days. (iv) The resulting milky suspensions were centrifuged; the centrifuge is a machine based on sedimentation principle that puts an object in rotation about an axis to separate particles of varying density. The precipitates were thoroughly washed with ethanol (CH_3–CH_2–OH) and dichloromethane (CH_2Cl_2). They were subsequently dried in air at 60°C.

Juttukonda et al. (2006) reported the synthesis of tin oxide nanoparticles stabilized by a variety of dendritic polymers (polymers having a branching treelike appearance). Their technique consists of the simple oxidation of an encapsulated stannate (containing SnO_3^{2-} units) salt via reaction with carbon dioxide (CO_2) under ambient environment. Synthesis of tin (IV) oxide nanoparticles was accomplished via the reaction of carbon dioxide with stannate ions immobilized by dendritic polymers. Stirring of an aqueous or ethanolic solution containing sodium stannate and a polymeric host was first carried out for a minimum period of 2 h to ensure saturated host–guest chelation (to firmly bind a metal ion with an organic molecule to form a ring-shaped molecular complex in which the metal is firmly bound and isolated; the resulting ring structure protects the mineral from entering into unwanted chemical reactions). Bubbling gaseous carbon dioxide through the room-temperature mixture with vigorous stirring for 30 min resulted in a light yellow solution.

The best molar ratio of dendrimer/sodium stannate [$Na_2SnO_3 \cdot 3H_2O$ or $Na_2Sn(OH)_6$], was 1:4. On introducing more stannate, a white cloudy precipitate formed through reaction with CO_2 owing to the presence of free SnO_3^{2-} ions in solution. Besides providing control over nanoparticle growth, their use of an undersupplied number of stannate ions resulted in the majority of primary amines left behind unchelated, accessible for subsequent reaction/functionalization, if wanted.

Tabatabaei et al. (2006) synthesized silica nanoparticles from tetraethylorthosilicate (TEOS):$SiC_8H_{20}O_4$, ethanol (C_2H_5OH), and deionized water. Colloidal silica nanoparticles with a constricted particle size distribution were obtained by the hydrolysis reaction of TEOS in ethanol containing water and ammonia (NH_3).

Materials: TEOS ($SiC_8H_{20}O_4$), NH_4OH solution, and ethanol (EtOH).

Method: (i) Reagents were mixed into two starting time solutions of ethanol: (a) TEOS/EtOH and (b) NH_4OH/H_2O/EtOH. By regulating the contents of solutions (a) and (b), the concentrations of TEOS, H_2O, and NH_4OH were fixed at prearranged values. The solutions were prepared in a glove box at room temperature under a controlled humidity of a few percentages. (ii) Solutions (a) and (b) were mixed with each other at 298 K. The mixture was subjected to vigorous stirring by hand for 6 s. Glycerol [$C_3H_5(OH)_3$] was added to the NH_4OH/H_2O/EtOH mixture. This was followed by addition of TEOS.

Condensation reaction began 2–10 min after the addition of TEOS; it is a reaction in which two molecules react with the resulting loss of a molecule of water (or other small molecule): the formal reverse of hydrolysis. The condensation reaction was easily observed. The hydrolysis reaction forming silicic acid (H_4O_4Si) is not seen. After this unseen reaction, the condensation of supersaturated silicic acid was indicated by an increasing dullness and opacity of the mixture. Then a turbid white suspension was formed.

3.2.5 Synthesis of Carbon Nanotubes

Presently, there are three foremost techniques to produce high-quality single-walled carbon nanotubes (SWCNTs): (i) laser ablation (the process of evaporating material from a target solid surface by irradiating it with pulses of laser beam), (ii) electric arc discharge (a self-maintaining direct-current electrical current flow through a gas, producing an ongoing, brightly glowing plasma discharge and characterized by a voltage drop approximately equal to the ionization potential of the gas), and (iii) chemical vapor deposition (CVD). *What are the salient features of these techniques?* Among the aforementioned techniques, laser ablation and arc discharge are physical methods. In fact, they are modified physical vapor deposition (PVD) techniques and involve the condensation of hot gaseous carbon atoms generated from the evaporation of solid carbon. However, the equipment requirements and the large amount of energy consumed by these methods make them mostly appropriate for laboratory work. The problem lies in scaling up these methods to fabricate large quantities of CNTs to cater to proliferating industrial applications. Both the laser ablation and arc discharge techniques suffer from limitation in the volume of sample they are able to produce in comparison to the size of the carbon source. In addition, larger percentages of impurities are found in the nanotubes synthesized by these methods in the

form of amorphous carbon and catalyst particles because of the high temperature nature of the heat source. Since the growth is difficult to control, the final product consists mostly of multi-walled carbon nanotubes (MWCNTs) with poor alignment. Nonetheless, the overall yield of CNTs produced is sufficient for laboratory-scale experiments and they have been mostly used for academic research.

What is the favorite technique for CNT manufacturing? It is CVD. *What is CVD?* CVD involves irreversible deposition of a solid from a gas or a mixture of gases through a heterogeneous chemical reaction. *Where does this reaction occur?* It takes place at the interface of gas–solid substrate. Depending on the deposition conditions, the growth process is restrained either by diffusion or by surface kinetics.

Generally speaking, CVD of hydrocarbons over a metal catalyst is a classical method that has been used by many people in the production of various carbon materials like carbon fibers and filaments. *What are the merits of CVD in general?* Its advantages are as follows: (i) CVD is the preferred technique for fabrication of thin layers of metals, insulators, and semiconductors on different substrates. (ii) It offers better growth controllability because of the equilibrium nature of the chemical reactions involved.

CVD of CNT uses carbon precursor gases such as methane (CH_4), ethylene (C_2H_4), acetylene (C_2H_2), benzene (C_6H_6), carbon monoxide (CO), and ethanol (C_2H_5OH). The process usually involves high-temperature decomposition of hydrocarbons in hydrogen over the catalyst, which is pre-deposited on the solid substrate. *What are the merits of CVD for CNTs?* Advantages of CVD for CNTs are as follows: (i) CVD has the capability to maneuver the size, shape, and alignment of the nanotubes through a cautiously designed patterning of the catalysts on the surface of the substrate. (ii) It is a continuous process and currently the best-known technique for high yield, and low impurity production of CNT at moderate temperatures. (iii) Amongst CNT production techniques, this method is easily up-scalable to industrial production. It is widely recognized as having the capability for large-scale production of CNTs, at least 10,000 tons/annum per plant (MacKenzie et al. 2008). In particular, the fluidized-bed CVD (FBCVD) technique (where the CVD reaction occurs within a fluidized bed of catalyst particles) offers the greatest expectation to produce high-quality CNTs, economically in large quantities (See and Harris 2007). A *fluidized bed* is formed when a quantity of a solid particulate substance, generally present in a holding vessel, is placed under appropriate conditions to cause the solid/fluid mixture to behave as a fluid. This is usually achieved by the introduction of pressurized fluid through the particulate medium, resulting in the medium acquiring many properties and characteristics of normal fluids, for example, the ability to free-flow under gravity, or to be pumped using fluid-type technologies.

Consequently, CVD has emerged as the most viable, commercial approach for manufacturing carbon nanotubes.

3.2.5.1 Arc Discharge Method of CNT Production

Arc discharge, initially used to produce C_{60} fullerenes (molecules composed entirely of carbon, in the form of a hollow sphere, ellipsoid, or tube; the third allotropic form of carbon material after graphite and diamond), creates CNTs through arc vaporization of two carbon rods placed end to end, separated by <1 mm, in an enclosure that is usually filled with inert gas (helium, He, or argon, Ar) at low pressure (Figure 3.2). A direct current of 50–100 A, driven by a potential difference ~30 V, creates a high-temperature (>3000°C) plasma discharge between the two electrodes. The discharge vaporizes the surface of one of the carbon electrodes and forms a small rod-shaped deposit on the other electrode.

What are the shortcomings of arc discharge method? Drawbacks of arc discharge method are as follows: (i) The inherent design of these systems poses limitations to the large-scale production of CNTs, for example, vacuum conditions are required to prevent the interference of formation of unwanted ions at the high temperatures used in arc discharge. But vacuum conditions are troublesome and expensive to scale up. (ii) The graphite targets and electrodes require continuous replacement as the synthesis proceeds, and, hence, these techniques cannot be operated continuously.

3.2.5.2 Laser Ablation Method of CNT Production

A laser source is used to generate high temperatures on a carbon target (Figure 3.3). The vaporized carbon rapidly cools in a carrier gas stream, for example, helium (He), and forms CNTs and other carbonaceous (consisting of, containing, relating to, or yielding carbon) secondary products. The use of two laser pulses in succession reduces the amount of carbon deposited as soot (a fine black or brown powder consisting of impure carbon particles that forms through incomplete combustion of a hydrocarbon) to the minimum extent. *How does this happen?* The second laser pulse breaks up the larger particles ablated (eroded by melting) by the first one, and feeds them into the maturing nanotubes structure. On the whole, the CNT material produced
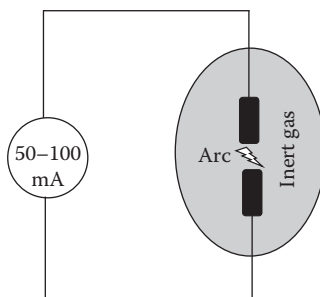
FIGURE 3.2
Arc discharge arrangement for CNT synthesis. (After Singh, P. et al., *J. Optoelectron. Biomed. M.,* 2, 91, 2010.)
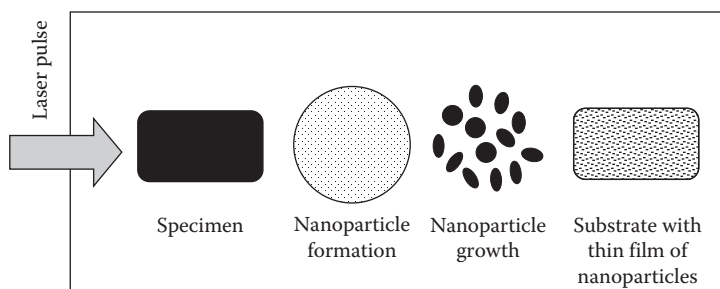
FIGURE 3.3
Laser ablation method of CNT production. (After Singh, P. et al., *J. Optoelectron. Biomed. M.*, 2, 91, 2010.)

by this method appears as an entangled mat of "stout cords," 10–20 nm in diameter and up to 100 µm or more in length.

3.2.5.3 Chemical Vapor Deposition Method of CNT Production

Large amounts of CNTs can be formed by catalytic CVD of acetylene (C_2H_2) over cobalt (Co) and iron (Fe) catalysts residing on silica (SiO_2) or zeolite (microporous, aluminosilicate minerals used as commercial adsorbents) support (Singh et al. 2010). High yields of SWCNTs have been obtained by catalytic decomposition of H_2/CH_4 mixture over well-dispersed metal particles such as cobalt (Co), nickel (Ni), and iron (Fe) on magnesium oxide (MgO) at 1000°C (Figure 3.4). The reduction produces very small transition metal (an element whose atom has an incomplete *d* sub-shell, or which can give rise to cations with an incomplete *d* sub-shell) particles at a temperature of usually >800°C. The disintegration of CH_4 over the newly formed nanoparticles hinders their further growth and thus results in a very high proportion of SWCNTs and few MWCNTs.

Figure 3.5 depicts how the synthesis of CNT takes place. In principle, CVD of CNT is a thermal dehydrogenation (a chemical reaction that involves the elimination of hydrogen; it is the reverse process of hydrogenation) reaction by which a transition metal catalyst, for example, iron, nickel, and cobalt, is used to bring down the temperature required in order to "crack" a gaseous hydrocarbon feed into carbon and hydrogen. In several ways, it is comparable to the large-scale synthesis of hydrocarbon fuels, for example, octane (chemical formula C_8H_{18}, and the structural formula $CH_3(CH_2)_6CH_3$), by catalytic cracking meaning the decomposition of heavy oils to produce lighter hydrocarbons. In this field, it has already been successfully scaled up to large throughputs; hence promising.

There exist a large number of CVD modalities. CVD variants include (i) fixed beds, (ii) fluidized beds, (iii) aerosols (colloidal particles dispersed in a gas), (iv) floating catalysts, and (v) combination methods, for example, plasma-enhanced (PECVD) and laser-assisted techniques, all of which are
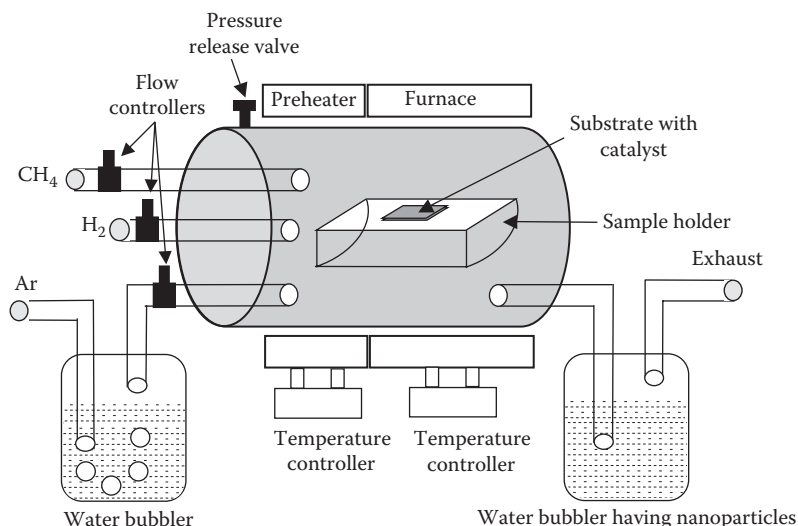
FIGURE 3.4
Chemical vapor deposition of carbon nanotubes: source of carbon atoms is usually an organic compound, the precursor (methane, ethylene, acetylene, benzene, carbon monoxide, or ethanol). The bubbler provides water vapor carried by argon gas into the reactor. Water serves as an oxidizing agent helping to remove amorphous carbon from the surface of the catalyst particle. In addition, hydrogen and argon are added to the gas mixture for better control of the carbon concentration in the gas phase. A sample holder accommodates the substrate with patterned catalysts on its surface. Transition metals such as Fe, Co, and Ni are frequently used as catalysts. Furnace temperature ranges between 600°C and 1200°C. Carbon nanotubes are deposited on the reactor wall usually constructed of quartz (http://classes.engr.oregonstate.edu/cbee/winter2005/che414/Honors/Bestwick-Nanotubes.ppt).

capable of producing freestanding (not supported by or adjoining another structure) CNTs. These freestanding CNTs resemble those grown using laser ablation or arc discharge, that is, they are not grown on templates (patterns or gauges, usually as thin boards/metal pieces, for guiding cutting or drilling). PECVD reactor is a commonly used equipment in microelectronics industry. Figure 3.6 shows the PECVD method for CNT formation.

The fixed-bed reactor has traditionally been used for CNT synthesis via CVD. A cross-flow setup is usually employed inside a horizontal furnace. In the fluidized bed CVD (FBCVD) process, the CVD reaction occurs within a fluidized bed of catalyst particles. Freestanding CNTs required in applications such as composite materials and energy-storage devices are produced cost effectively by this technique.

3.2.5.4 Difficulties Faced with Carbon Nanotubes

What are the obstacles to CNT usage in nanosensor fabrication? One major shortcoming of conventional carbon nanotube device fabrication techniques is the inability to scale up the processes for fabrication of a large number of
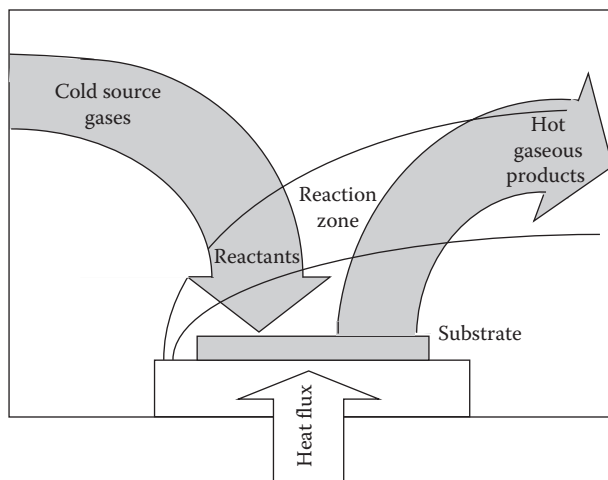
FIGURE 3.5
Mechanism of CVD of CNTs: gases enter chamber at room temperature (cooler than the reaction temperature). They are heated as they approach the substrate. They react with the substrate or undergo chemical reaction in the "reaction zone" before reacting with the substrate forming CNT deposits. The gaseous products are then removed from the reaction chamber (http://classes.engr.oregonstate.edu/cbee/winter2005/che414/Honors/Bestwick-Nanotubes.ppt).
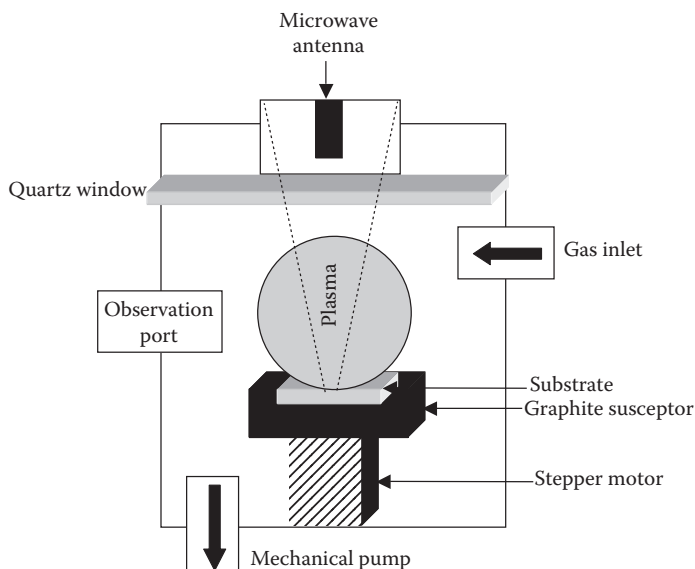
FIGURE 3.6
Plasma-enhanced CVD-based CNT synthesis: alumina substrate (Al_2O_3) is coated with ferric nitrate [$Fe(NO_3)_3 \cdot 6H_2O$] (the catalyst). Methane (CH_4) and hydrogen (H_2) are introduced into the vacuum chamber at a total pressure of 15 Torr, and the chamber is then heated to 850°C–900°C with an input microwave power of 600 W. CNTs are grown from the substrate material (http://classes.engr.oregonstate.edu/cbee/winter2005/che414/Honors/Bestwick-Nanotubes.ppt).

devices on a single chip. It may be carefully noted that increasing the proportion of nanosensor device fabrication process using CNTs is being referred to, not the CNT production process itself, which was the focus of attention in the previous discussion. For this purpose, accurate assembly of single-nanotube devices with an integration density of several million devices per square centimeter is mandatory. Therefore, for CNT sensors to become a reality, it is very important to scale up the fabrication technique to simultaneously and reproducibly fabricate a very large number of such devices on a single chip, each accessible separately for electronic transport. Conventional nanotube growth and device fabrication techniques using CVD or spin casting have failed to achieve this goal because of lack of precise control over nanotube positioning and orientation. Due to the difficulties encountered in correctly handling and manipulating these nanoscale objects at the individual level, various attempts to assemble them into functional sensor devices have, unfortunately, been successful only to a limited extent. In the ideal case, it should be possible to "pick and place" an individual nanotube at a predefined location and orientation, forming robust, low-resistance, ohmic contacts to two metallic leads. Furthermore, it should be possible to do this at a scalable integration density with each nanotube forming a separately addressable nanosensor.

Do CNTs possess toxicity? Yes. In the context of toxicology, the nanometer-scale dimensions of CNT make quantities of milligrams possess a large number of cylindrical, fiber-like particles with a concurrent very high total surface area. Batches of pristine CNT (non-purified and/or nonfunctionalized), after synthesis, contain impurities such as amorphous carbon and metallic nanoparticles (catalysts: Co, Fe, Ni, and Mo), which are the sources of severe toxic effects. Platelet aggregation is induced by both SWCNTs and MWCNTs but not by the C_{60} fullerenes that are used as building blocks for these CNTs. Platelets are the small, colorless, disc-shaped fragments having no nucleus, found in blood and involved in its clotting.

3.3 Micro- and Nanoelectronics Division

3.3.1 Semiconductor Clean Room

A clean room is an enclosed clean space in which semiconductor manufacturing takes place. Airborne particles are reduced to technically feasible minimum. The clean room is defined as a room with air containing N particles per cubic foot of size $\leq 0.5\,\mu m$ where $N \leq 1$, 10, 100, 1,000, 10,000, etc.; accordingly it is designated as Class 1, Class 10, Class 100, Class 1000, and Class 10000. *What are the temperature and humidity guidelines for semiconductor fabrication clean rooms?* Temperature (20°C–25°C) and humidity (40%–50%) of ambient air are strictly controlled, both for efficient running

of equipment and to maintain the critical ambience requirements for carrying out the fabrication processes effectively. Noncompliance to these conditions will invariably lead to equipment breakdown and device processing faults.

3.3.2 Silicon Single Crystal Growth and Wafer Production

What is the starting material for integrated circuit fabrication? Integrated circuits are built on single-crystal silicon substrates that possess a high level of purity and crystalline perfection. *How is such high-grade Si material obtained*? The acquisition of such high-grade starting silicon material involves two major steps: (i) refinement of raw material (such as quartzite, a hard metamorphic rock that was originally sandstone) into electronic-grade polycrystalline silicon (EGS) using a complex multistage process and (ii) growing of single-crystal silicon from this EGS either by Czochralski (C-Z) or Float Zone (F-Z) process. Silicon is readily available through the treatment of silica, SiO_2, with pure graphite (as coke) in an electric furnace.

$$SiO_2 + 2C \rightarrow Si + 2CO \tag{3.2}$$

Very pure silicon is made by the reaction of $SiCl_4$ with hydrogen, followed by zone refining (also known as zone purification, a technique to purify materials in which a narrow molten zone is moved slowly along the complete length of the specimen to cause impurity segregation, and which depends on the differences in composition of the liquid and solid in equilibrium) of the resultant silicon. Material used in single crystal growth is EGS (99.999999999% pure).

C-Z crystal growth involves the crystalline solidification of atoms from a liquid phase at an interface. The silicon charge, taken in a fused silica crucible, is melted (Si melting point = 1421°C), and a seed of crystal silicon with precise orientation is introduced into the molten silicon. The seed crystal is withdrawn at a controlled rate, and the seed crystal and the crucible are rotated in opposite directions. By controlling the temperature gradients, rate of pulling and speed of rotation, a large, single-crystal, cylindrical ingot (a material, usually metal that is cast into a shape suitable for further processing) is extracted from the melt.

The F-Z process entails the passing of a molten zone through a polysilicon rod that approximately has the same dimensions as the final ingot. A polycrystalline rod is passed through an RF heating coil, creating a localized molten zone from which the crystal ingot is grown. A seed crystal (small piece of single crystal material from which a large crystal of the same material typically is grown) is used at one end to initiate the growth process. The purity of an ingot produced by the F-Z process is higher than one produced by the C-Z process. These impurities are mainly introduced by the material of the crucible (a refractory container used for metal, glass, etc., production),

which holds the silicon melt. As such, C-Z silicon has much higher oxygen and carbon impurities than F-Z silicon.

What are the typical operating parameters of C-Z and F-Z processes? In a representative C-Z process, induction heating (heating an electrically conducting object usually a metal, by electromagnetic induction) at 0.4 MHz melts the Si held in a stationary clear fused quartz crucible nested in a graphite susceptor (material used for its ability to absorb electromagnetic energy and convert it to heat). The ambient is argon at 0.1 bar above atmospheric pressure. Typically, seed crystal is ~0.5 cm in diameter and 10 cm long. The growth rate is 1.5 mm min^{-1} with a crystal rotation rate of 14 RPM. F-Z growth is carried out on Si using RF heating with a stationary one-turn coil operating at 2 MHz in an argon ambient at 0.3 bar above atmospheric pressure, to minimize heater or crucible sources of O, C, and other impurities. The typical growth rate is 3 mm min^{-1} with a 13–16 RPM crystal rotation rate and a 2–3 RPM feed rod rotation rate (crystal on the bottom side, moving downward).

3.3.3 Molecular Beam Epitaxy

Epitaxy or *epitaxial growth* is the process of depositing a thin layer (0.5–20 μm) of single crystal material over a monocrystalline substrate. Molecular beam epitaxy (MBE) is a technique for epitaxial growth via the interaction of one or several molecular or atomic beams that occurs on the surface of a heated crystalline substrate. The solid source materials are housed in evaporation cells to provide an angular distribution of atoms or molecules in a beam. The substrate is heated to the necessary temperature. When needed, it is continuously rotated to improve the homogeneity of growth. The molecular beam condition that the mean free path λ (the average distance covered by a moving particle between successive impacts) of the particles should be larger than the geometrical size of the chamber is easily fulfilled if the total pressure does not exceed 10^{-5} Torr. Also, the condition for growing a sufficiently clean epilayer must be satisfied. It is necessary that the monolayer deposition times of the beams t_b and the background residual vapor t_{res}, should obey the relation $t_{res} < 10^{-5} t_b$. Thus, MBE is a technique of growing single crystals in which beams of atoms or molecules bombard a single-crystalline substrate in a vacuum, giving birth to crystals whose crystallographic orientation is related to that of the substrate. In MBE, a source material is heated to produce an evaporated beam of particles. These particles propagate and disseminate through a very high vacuum (10^{-8} Pa) to the substrate, where they condense. MBE has lower throughput than other forms of epitaxy.

3.3.4 Mask Making

Based on the electrical specifications of the given circuit, a geometrical layout for the same is made defining the areas for diffusion windows, contact

window openings, metal patterns, etc., for resistors, capacitors, diodes, transistors, and other devices. Then, in accordance with the fabrication process sequence, a series of masks in the form of aligned patterns are designed. Mask making consists of two steps: (i) layout generation (defining the pattern that will appear on the mask) and (ii) pattern transferring to the mask.

Layout is the process of defining the patterns that will be transferred to the masks, and as such will define the geometry of devices. Layout is typically performed in a graphical editing tool such as LEDIT, Cadence, etc. Distinct layer names are used for each layer in the layout; the layer designation for each layer should be remembered. The polarity of the mask must be indicated by either bright-field (space left open in layout is transparent) or dark-field (space left open in layout is opaque).

For optical lithography, commonly used substrates are soda-lime glass and quartz. Quartz masks have a matching coefficient of thermal expansion to Si, but are more expensive than soda-lime masks. The two most widespread mask coating materials used for optical lithography are emulsion and chrome. Photographic emulsion is a light-sensitive colloid, such as silver halide crystals (silver bromide AgBr, silver chloride [AgCl] and silver iodide [AgI]) suspended in gelatin (a protein substance derived from collagen, a natural protein present in the tendons, ligaments, and tissues of mammals). The light-exposed crystals are reduced by the developer to black metallic Ag particles forming the image. Emulsion is much cheaper to make; however, the coating is not of as high a quality as chrome, and does not forebear long the rigors of processing.

For the production of chromium film masks, a thin layer of chromium (Cr), ~0.1 μm thick, is deposited onto a transparent substrate, most preferably glass, by vacuum deposition. The formed chromium layer is then selectively etched into the desired pattern by photoetching process, which comprises the following steps: (i) applying a photoresist material, typically a natural or synthetic polymer, to the layer to be etched; (ii) selectively exposing the photoresist whereby certain portions thereof become soluble or insoluble upon development; (iii) removing the soluble portions that yield a negative or positive of the desired pattern; and (iv) etching the thus exposed portions of the underlying layer.

The exposure of the resist is commonly done in two ways: (i) laser pattern generation and (ii) electron-beam (E-beam). Laser writing can create features down to 500 nm whereas E-beam reaches up to 10 nm. Laser pattern generation involves the use of a rectangular shutter, which is capable of variable size and orientation. The shutter is used to sequentially expose the resist on the mask substrate to a laser beam until the requisite pattern is fully defined. As the mask is exposed sequentially, small intricate features and curved features may require a large number of flashes to define the pattern. A pattern with a high flash count may make this method of mask pattern transfer uneconomic.

E-beam is a technique, in which an electron beam is used to selectively expose the resist on the mask. The beam is raster scanned across the wafer, defining the pattern. The beam may be made very narrow, so that very intricate features may be defined using this approach. The narrower the beam used, the longer the write time for a mask (as more scans are required), which increases the price.

3.3.5 Thermal Oxidation

It is a method to produce a thin layer of oxide (usually silicon dioxide) on the surface of a silicon wafer by exposing the wafer to an oxidizing environment of O_2 or H_2O at elevated temperature (700°C–1200°C) (Figure 3.7); accordingly the process is called dry and wet oxidation. Silicon is consumed in the process. Oxidation occurs at the silicon–silicon dioxide interface.

$$Si\,(solid) + O_2\,(gas) = SiO_2\,(solid);$$

$$Si\,(solid) + 2H_2O\,(gas) = SiO_2\,(solid) + 2\,H_2\,(gas) \qquad (3.3)$$
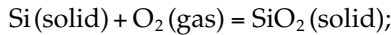
This oxide layer is (i) an excellent insulator and (ii) serves as a mask against diffusion of impurities like phosphorous or boron into silicon.

Deal–Grove model is a model describing kinetics of thermal oxidation of silicon based on chemical reaction between silicon and oxidizing species. It is most accurate for oxides thicker than about 30 nm but of limited use for oxide thinner than about 10 nm. It assumes surface reaction–controlled oxide growth in early stage of oxidation (linear regime) and controlled by diffusion of oxidizing species through the oxide during extended oxidation (parabolic
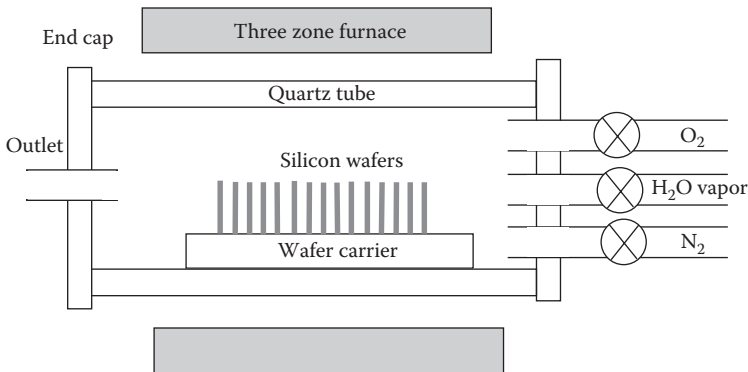
FIGURE 3.7
Typical thermal oxidation furnace for silicon wafers.

regime). It is one among the best established, celebrated models in silicon processing.

According to the Deal–Grove model, the thickness of thermal oxide is expressed as

$$t_{ox} = \left(\frac{A}{2}\right)\left\{\sqrt{1 + \frac{t+\tau}{(A^2/4B)}} - 1\right\} \tag{3.4}$$

where

B/A is linear rate constant

B is parabolic rate constant

τ denotes a shift in the time coordinate to account for the presence of the initial oxide layer

Example 3.1

Wet oxidation of silicon is carried out at 1000°C. Given the following values of parameters in the Deal–Grove model: $A = 0.226\,\mu m$, $B = 0.287\,\mu m^2\ h^{-1}$, and $\tau = 0$, find the oxide thickness for growth times $t = 0.01\,h$ and $t = 100\,h$.

In Equation 3.4, $A = 0.226\,\mu m$, $B = 0.287\,\mu m^2\ h^{-1}$, and $\tau = 0$.

For $t = 0.01\,h$, we have

$$t_{ox} = \left(\frac{A}{2}\right)\left\{\sqrt{1 + \frac{t+\tau}{(A^2/4B)}} - 1\right\}$$

$$= \left(\frac{0.226 \times 10^{-4}\ cm}{2}\right)$$

$$\times \left[\sqrt{1 + \frac{0.01\,h + 0}{\left\{(0.226 \times 10^{-4}\ cm)^2 / (4 \times 0.287 \times 10^{-4} \times 10^{-4}\ cm \times cm/h)\right\}}} - 1\right] \tag{3.5}$$

$$= 1.13 \times 10^{-5}\left[\sqrt{1 + \frac{0.01}{\left\{5.1076 \times 10^{-10}/1.148 \times 10^{-8}\right\}}} - 1\right]$$

$$= 1.13 \times 10^{-5}\left[\sqrt{1 + \frac{0.01}{4.449 \times 10^{-2}}} - 1\right]$$

$$= 1.13 \times 10^{-5}\left[\sqrt{1.2248} - 1\right] = 1.206 \times 10^{-6}\ cm = 0.01206\ \mu m \tag{3.6}$$

For $t = 100\,h$, we get

$$t_{ox} = \left(\frac{A}{2}\right)\left\{\sqrt{1 + \frac{t + \tau}{\left(A^2/4B\right)}} - 1\right\}$$

$$= \left(\frac{0.226 \times 10^{-4}\ cm}{2}\right)$$

$$\times \left[\sqrt{1 + \frac{100\ h + 0}{\left\{\left(0.226 \times 10^{-4}\ cm\right)^2 \Big/ \left(4 \times 0.287 \times 10^{-4} \times 10^{-4} cm \times cm/h\right)\right\}}} - 1\right]$$

(3.7)

$$= 1.13 \times 10^{-5}\left[\sqrt{1 + \frac{100}{\left\{5.1076 \times 10^{-10}\big/1.148 \times 10^{-8}\right\}}} - 1\right]$$

$$= 1.13 \times 10^{-5}\left[\sqrt{1 + \frac{100}{4.449 \times 10^{-2}}} - 1\right]$$

$$= 1.13 \times 10^{-5}\left[\sqrt{2.2487 \times 10^{3}} - 1\right] = 5.2455 \times 10^{-4}\ cm = 5.2455\ \mu m$$

(3.8)

As the thickness of gate dielectric scales below 2 nm, leakage current due to tunneling rises drastically, leading to heavy power consumption. Materials with high dielectric constants such as silicon oxynitride (SiON), hafnium dioxide (HfO_2), zirconium dioxide (ZrO_2), and titanium dioxide (TiO_2) permit increasing the thickness of the gate dielectric, thereby alleviating the tunneling current problem. Thus, substituting the silicon dioxide gate dielectric with a high-κ material yields the required high gate capacitance without the attendant undesirable leakage current.

3.3.6 Diffusion of Impurities in a Semiconductor

The process by which molecules spread from areas of high concentration to areas of low concentration is used for junction formation, that is, transition from P- to N-type or vice versa. It is typically accomplished by the process of diffusing the appropriate dopant impurities in a high-temperature furnace. At high temperature, many atoms in the semiconductor move out of their lattice site, leaving vacancies into which impurity atoms move to take up their places. The impurities, thus, diffuse by this type of vacancy motion and occupy lattice position in the crystal after it is cooled.

The behavior of diffusion particles is governed by Fick's laws, which, when solved for appropriate boundary conditions, give various dopant distributions

called *profiles*. These profiles are approximated during actual diffusion processes. *Fick's first law* states that the flux, *J*, of a component of concentration, *C*, across a membrane of unit area, in a predefined plane, is proportional to the concentration differential across that plane; *Fick's second law* states that the rate of change of concentration in a volume element of a membrane, within the diffusional field, is proportional to the rate of change of concentration gradient at that point in the field. Whereas Fick's first law allows the diffusive flux *J(x)* to be calculated as a function of concentration gradient, the second law enables the concentration function *C(x)* to be calculated as a function of time.

Fick's laws are applied to solve diffusion problems in semiconductor fabrication. Depending on boundary conditions, two types of solutions are obtained. These solutions provide two types of impurity distribution: (i) *constant or infinite source distribution* following *complementary error function (erfc)* and (ii) *limited or finite source distribution* following *Gaussian distribution function*. Mathematically, the constant source diffusion is described by the equation

$$N(x,t) = N_s \mathrm{erfc}\left(\frac{x}{2\sqrt{Dt}}\right) \qquad (3.9)$$

where
$N(x,t)$ is the concentration of impurity in atoms cm^{-3} at distance x at time t
N_s is the surface concentration
D is the diffusion constant of the impurity at the temperature of diffusion

For limited source diffusion, the equation is

$$N(x,t) = \frac{Q}{\sqrt{\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right) \qquad (3.10)$$

Q is the total impurity in atoms cm^{-2} given by

$$Q = N_s \sqrt{\pi Dt} \qquad (3.11)$$

How is diffusion performed in practice? Practically, diffusion of impurities in a semiconductor is performed as a sequence of two steps. The two-step diffusion consists of a deposition step and a drive-in step. The unlimited source diffusion is called the impurity *pre-deposition step*. The limited source diffusion is known as the impurity *drive-in step*.

Example 3.2

Calculate the junction depth for a constant-source Boron diffusion into an N-type silicon substrate (1×10^{15} cm^{-3}) performed at 1200°C for 100 min. Given that solid solubility of boron at this temperature $= 2.3 \times 10^{20}$

atoms cm^{-3}, diffusion coefficient $= 2.5 \times 10^{-12}$ cm^2 s^{-1}, and $erfc^{-1}$ (0.999 9957) = 3.25.

The required junction depth is obtained by applying the equation

$$N(x,t) = N_s \text{erfc}\left(\frac{x}{2\sqrt{Dt}}\right) \qquad (3.12)$$

from which

$$1 \times 10^{15} = (2.3 \times 10^{20})\text{erfc}\left\{\frac{x}{2\sqrt{2.5 \times 10^{-12} \times 100 \times 60}}\right\} \qquad (3.13)$$

so that

$$\frac{1 \times 10^{15}}{2.3 \times 10^{20}} = \text{erfc}\left\{\frac{x}{2\sqrt{2.5 \times 10^{-12} \times 100 \times 60}}\right\} \qquad (3.14)$$

or

$$4.3478261 \times 10^{-6} = \text{erfc}\left\{\frac{x}{2.45 \times 10^{-4}}\right\} \qquad (3.15)$$

or

$$\text{erfc}^{-1}\left(1 - 4.3478261 \times 10^{-6}\right) = \frac{x}{2.45 \times 10^{-4}} \qquad (3.16)$$

or

$$x = 2.45 \times 10^{-4} \times \text{erfc}^{-1}\left(1 - 4.3478261 \times 10^{-6}\right)$$

$$= 2.45 \times 10^{-4} \times \text{erfc}^{-1}\left(0.99999565\right)$$

$$= 2.45 \times 10^{-4} \times 3.25 = 7.96 \times 10^{-4} \text{ cm} = 7.96 \text{ μm} \qquad (3.17)$$

Example 3.3

For fabricating a PN junction, P-type diffusion was done in an N-type silicon wafer of background concentration 1×10^{15} atoms cm^{-3}. The final impurity profile was of the form exp $\{-x^2/(4Dt)\}$ where the junction depth (x) was 2 μm and the time (t) was 1 h. A surface concentration of

1×10^{18} atoms cm^{-3} was achieved. Find the diffusivity of the dopant and the pre-deposition dose.

From the equation

$$Q = N_s \sqrt{\pi D t} \tag{3.18}$$

we get

$$Q = N_s \sqrt{\pi D t} = 1 \times 10^{18} \times \sqrt{3.14 \times D \times 1 \times 60 \times 60}$$

$$= 1 \times 10^{18} \times \sqrt{3.14 \times D \times 1 \times 60 \times 60} = 1 \times 10^{18} \times \sqrt{11,304D} \tag{3.19}$$

Applying the equation

$$N(x,t) = \frac{Q}{\sqrt{\pi D t}} \exp\left(-\frac{x^2}{4Dt}\right) \tag{3.20}$$

we have

$$1 \times 10^{15} = \frac{1 \times 10^{18} \times \sqrt{11,304D}}{\sqrt{3.14 \times D \times 1 \times 60 \times 60}} \exp\left\{-\frac{\left(2 \times 10^{-4}\right)^2}{4D \times 1 \times 60 \times 60}\right\}$$

$$= \frac{1 \times 10^{18} \times \sqrt{11,304D}}{\sqrt{11,304D}} \exp\left\{-\frac{4 \times 10^{-8}}{1.44 \times 10^4 D}\right\} \tag{3.21}$$

yielding

$$\frac{1 \times 10^{15}}{1 \times 10^{18}} = \frac{\sqrt{11,304D}}{\sqrt{11,304D}} \exp\left\{-\frac{4 \times 10^{-8}}{1.44 \times 10^4 D}\right\} \tag{3.22}$$

or

$$10^{-3} = \exp\left\{-\frac{4 \times 10^{-8}}{1.44 \times 10^4 D}\right\} = \exp\left(-\frac{2.78 \times 10^{-12}}{D}\right) \tag{3.23}$$

Taking natural logarithm of both sides

$$\ln(10^{-3}) = -\frac{2.78 \times 10^{-12}}{D} \tag{3.24}$$

giving

$$D = -\frac{2.78 \times 10^{-12}}{\ln(10^{-3})} = -\frac{2.78 \times 10^{-12}}{-6.907755} = 4.0245 \times 10^{-13} \text{ cm}^2 \text{ s}^{-1} \quad (3.25)$$

Substituting for D from Equation 3.25 in Equation 3.19

$$Q = 1 \times 10^{18} \times \sqrt{11,304D} = 1 \times 10^{18} \times \sqrt{11,304 \times 4.0245 \times 10^{-13}}$$

$$= 6.745 \times 10^{13} \text{ atoms cm}^{-2} \quad (3.26)$$

For phosphorus diffusion, compounds such as PH_3 (phosphine) and $POCl_3$ (phosphorus oxychloride) are used. In the case of P-diffusion using $POCl_3$, Figure 3.8, the reactions occurring at the silicon wafer surface are

$$Si + O_2 = SiO_2 \text{ (silica glass)}; \quad 4POCl_3 + 3O_2 = 2P_2O_5 + 6Cl_2 \quad (3.27)$$

This results in the production of a glassy layer on the silicon wafers, a mixture of phosphorus glass and silica glass called phosphosilicate glass (PSG), which is a viscous liquid at the diffusion temperatures. The mobility (ease of motion) of the phosphorus atoms in this glassy layer and the phosphorus concentration is such that the phosphorus concentration at the silicon surface is maintained at the solid solubility limit throughout the time of the diffusion process. Solid solubility is the maximum concentration of impurity that can dissolve in silicon under
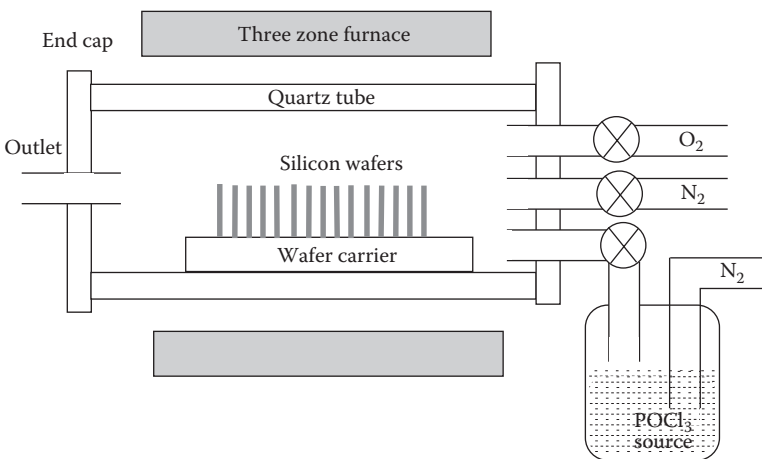
FIGURE 3.8
Thermal diffusion furnace for N-type phosphorous doping in silicon wafers using phosphorous oxychloride source.

a given temperature and pressure. This diffusion step is referred to as pre-deposition step in which the dopant atoms deposit into the surface regions (~0.1 µm depth) of the silicon wafers. *What is the use of PSG layer?* The PSG is preferable because it protects the silicon atoms from pitting or evaporating and acts as a "getter" (a material added in small amounts during a chemical reaction to absorb impurities and contaminants). It is etched off before next diffusion.

The pre-deposition step is followed by a second diffusion process in which the external dopant source (PSG) is removed such that no additional dopants enter the silicon. During this diffusion process, the dopants that are already present in the silicon move further inside and are thus redistributed. The junction depth increases, and at the same time the surface concentration decreases. This type of diffusion is called *drive-in*.

For boron diffusion, a controlled flow of carrier gas (N_2) is bubbled through boron tribromide, which with oxygen again produces boron trioxide (BSG) at the surface of the wafers as per the following reaction:

$$4BBr_3 + 3O_2 = B_2O_3 + 2Br_2 \tag{3.28}$$
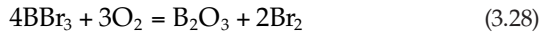
The result is deposition of a glassy layer on the silicon surface, which is a mixture of silica glass (SiO_2) and boron glass (B_2O_3) called borosilicate glass (BSG). The BSG glassy layer is a viscous liquid at the diffusion temperatures and the boron atoms move around relatively easily. The rest of process details for boron diffusion are similar to phosphorous diffusion.

3.3.7 Ion Implantation

Semiconductor manufacturers today use ion implantation for almost all doping in silicon ICs. *What are the common implanted species?* The most commonly implanted species are arsenic (As), phosphorus (P), boron (B), indium (In), antimony (Sb), germanium (Ge), nitrogen (N_2), hydrogen (H_2), and helium (He). Ion implantation works by ionizing the required atoms, accelerating them in the electric field, selecting the correct species by an analyzing magnet and bombarding the substrate with the ion beam in a pre-calculated manner.

Concentration and depth of the desired dopant are specified directly in the equipment settings for implant dose and energy, respectively. Therefore, the two key parameters defining the implant profile are (i) dose Φ (usually given in atoms cm^{-2}) and (ii) energy E (in keV). The dose is related to the beam current I. Typical beam currents and implantation doses range from 1 µA–30 mA and 10^{11}–10^{16} atoms cm^{-2}. The lowest energies used start at the sub keV area for ultrashallow junctions and ascend to the MeV range for deep wells.

What is the construction of an ion implanter? An ion implantation machine is constructed to deliver a beam of ions of a particular impurity at a specified energy and in a given dose to the surface of a silicon wafer. Figure 3.9 shows
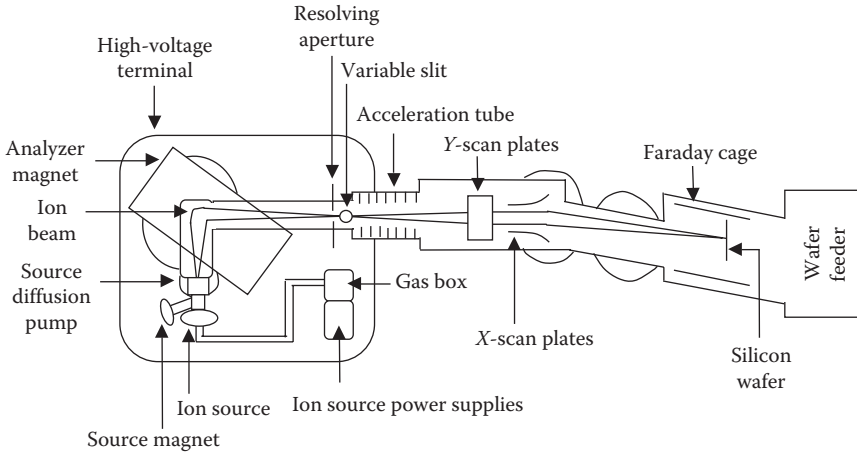
FIGURE 3.9
Parts of an ion-implantation machine.

the schematic diagram of an ion-implantation machine. A source of gas such as boron trifluoride (BF_3) feeds a small quantity of source gas into the ion source, in which a heated filament splits up the molecules into charged fragments, forming an ion-plasma containing the desired ions along with other ions and any accruing contamination. A high voltage ~20 kV extracts the ions and pushes them into the analyzer. Pressure in the remaining equipment is maintained at 10^{-6} Torr to avoid collisions and scattering of ions from the gas molecules. The magnetic field of the analyzer is carefully selected to allow the passage of ions of the required charge-to-mass ratio without blockage by the analyzer walls. Thus, the chosen ions of required charge-to-mass ratio continue their onward journey and enter the acceleration tube where they are accelerated to achieve the desired implantation energy as they move from high voltage to ground. The aperture collimates the ion beam. The *X*- and *Y*-electrostatic deflection plates are employed for scanning the surface of the wafer with the ion beam. A commercial ion implanter has typical dimensions of 6 m length × 3 m width × 2 m height.

3.3.8 Photolithography

Photolithography is the process of transferring geometric shapes or patterns on a mask to the surface of a silicon wafer coated with photoresist. *What is a photoresist?* It is a light-sensitive material performing two basic functions: (i) precise pattern formation and (ii) protection of the substrate from chemical attack during the etch process. Photolithography is the means by which the small-scale features of integrated circuits are created. The steps involved in the photolithographic process are wafer cleaning, barrier layer formation, photoresist application, soft baking (the step during which the solvents are removed

from the photoresist), mask alignment, exposure and development, and hard-baking (step to harden the photoresist and improve adhesion of the photoresist to the wafer surface). A diffusion barrier layer is a layer of thermally grown silicon dioxide that blocks the entry of dopant impurities like phosphorous and boron into the silicon wafer so that these impurities enter only through the windows in the oxide layer etched after the photolithographic operation.

What is the photolithographic process? A resist is applied to the surface using a spin-coating machine. This device holds the wafer of semiconductor, using a vacuum, and spins it at a high speed (3000–6000 RPM) for a period of 15–30 s. A small quantity of resist is dispensed in the center of the spinning wafer. The rotation causes the resist to be spread uniformly across the surface of the wafer with excess resist being thrown off by centrifugal action. Preparation of the resist is completed by a prebake step, where the wafer is gently heated in a convection oven and then a hotplate to evaporate the resist solvent, as already mentioned, and to partially solidify the resist.

The photomask (an opaque plate with transparent regions that allow light to shine through in a defined pattern) used in photolithography is created by a photographic process and developed onto a glass substrate. Chrome (slang for chromium) on quartz glass is used for the high-resolution deep UV (electromagnetic radiation between violet light and x-rays, having wavelengths of 200–400 nm) lithography. Depending on the design of the photolithography machine, there are three options: (i) the mask may be in contact with the surface, (ii) very close to the surface, or (iii) it is used to project the image of the pattern on to the surface of the substrate. These methods are called, not surprisingly, contact, proximity and projection, respectively. Figure 3.10 shows a schematic diagram of these methods. Option (i) provides good resolution but mask is spoilt, sometimes excessively damaged, on repeated abrasion with substrate. Option (ii) increases the life of mask but resolution is slightly sacrificed. Option (iii) is good for both the mask and the resolution, hence the best solution.

What happens during resist exposure? During the exposure process, the resist undergoes a chemical reaction. Depending on the chemical composition of resist, it can react in two ways when the light strikes the surface: (i) The action of light on a positive resist causes it to become soft and easily removable where it has been exposed to the light. (ii) A negative resist has the reverse property. Exposure to UV-light causes the resist to become hard and difficult to remove. After the developing process (dissolving exposed photoresist), a negative of the mask remains as a pattern of resist.

What is the purpose of post baking? The wafer undergoes a postbake process to further harden and remove any residue of the developer.

What mechanisms are involved in resist usage? A positive resist comprises two components: a resin (clear to translucent, solid or semisolid, viscous compound that can be hardened with treatment) and a photoactive compound (a dissolution inhibitor), dissolved in a solvent. On UV exposure, it becomes
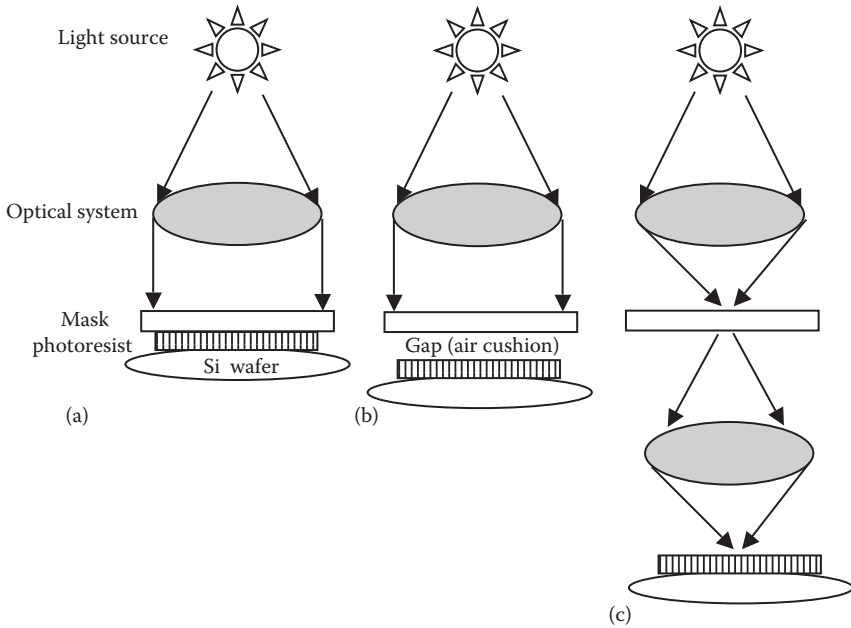
FIGURE 3.10
Photolithographic techniques: (a) contact, (b) proximity, and (c) projection.

more soluble in an aqueous developer solution. The negative photoresist also consists of two parts: a chemically inactive polyisoprene (monomer isoprene = C_5H_8) rubber, the film-forming component and a photoactive agent. On exposure to UV, the photoactive agent reacts with the rubber to form cross-links (bonds that link one polymer chain to another; they can be covalent bonds or ionic bonds) between the rubber molecules making the rubber less soluble in an organic developer solvent. Therefore, the developer solvent dissolves the unexposed resist whereas the exposed resist swells as the uncross-linked molecules are dissolved. This swelling causes distortion of the features in the pattern limiting the resolution to 2–3 times the initial film thickness.

Which resist provides better resolution, positive or negative, and why? Positive resist gives better resolution. Because in a positive photoresist the unexposed resist regions do not swell much in the developer solution, it is evident that a higher resolution is achievable with a positive photoresist.

3.3.8.1 Physical Limits

High spatial resolution needed in the image on the wafer necessitates large numerical aperture (NA) lenses. *Resolution* is defined as the ability of a microscope to allow the viewer to distinguish between small objects. It is

the shortest distance between two points on a specimen that can be distinguished by the observer or camera system as separate entities. The *Numerical Aperture* of a lens (a number that expresses the ability of a lens to resolve fine detail in an object being observed) is

$$NA = n \sin \theta \tag{3.29}$$

where
   n is the refractive index of the working medium
   θ is the half-angle of the maximum cone of light that can enter or exit the lens

Diffraction is the change in the directions and intensities of a group of waves after passing by an obstacle or through an aperture whose size is approximately the same as the wavelength of the waves. The *diffraction limit of resolution* on the wafer is

$$x_d = \frac{\eta \lambda}{NA} \tag{3.30}$$

where
   λ is the wavelength of the illuminating light for exposure
   η is a constant

It should be smaller than the minimum pattern dimension. Computerized lens designs have increased NA to values close to unity and appreciably reduced η, rendering the possibility of fabricating features with size considerably smaller than the exposing wavelength. However, using a high NA for increasing resolution results in a small depth of field (DoF: the distance between the nearest and farthest objects that appear in acceptably sharp focus). At a distance d from the focal plane, a point becomes an area of diameter

$$r = NA \times d \tag{3.31}$$

In lithography, *depth of field* implies the vertical distance from the focal plane at which r is equal to the diffraction limit, $r = x_d$. Thus,

$$DoF = \frac{\eta \lambda}{NA^2} \tag{3.32}$$

and the lithographic depth of focus decreases as the square of NA. DoF may be looked upon from an alternative viewpoint by eliminating NA from r and x_d, obtaining

$$\mathrm{DoF} = \frac{x_\mathrm{d}^2}{\eta\lambda} \qquad\qquad (3.33)$$

Hence, DoF is increased by using the smallest possible wavelength for a given diffraction resolution x_d.

3.3.8.2 Optical Lithography

The light sources for lithography have progressively used diminishing wavelengths to achieve the high resolution necessary for delineating small features. A new lithographic generation means that a different shorter wavelength light source, adequately intense to permit high throughput, must be resorted to. Furthermore, new resists optimized for this wavelength are sought. Lenses are redesigned and new transparent materials for lenses and mask supports are required. For 248 nm deep UV, KrF excimer laser (a form of ultraviolet laser; "excimer" is short for "excited dimer" and the dimmer is a molecule consisting of two identical simpler molecules) is the light source and the lens is made of fused quartz (glass containing primarily silica in amorphous form manufactured by melting naturally occurring quartz crystals of high purity) whereas for 13.4 nm extreme UV, stable, high power, gas discharge plasma is used as a light source and reflective optics is applied.

3.3.8.3 Electron-Beam Lithography

Exposure with a direct-writing scanning electron beam for producing line widths ~10–20 nm can help in overcoming the problem of light source. Additionally, it relieves the limits associated with resolution and depth of focus (the range of image distances corresponding to the range of object distances included in DoF). However, E-beam lithography (EBL) suffers from the following: (i) speed limitation and is therefore suited for low-volume production; (ii) also, the equipment is expensive; and (iii) as pattern generation is serial, for high throughput, parallelism is necessary by simultaneously exposing a considerable area instead of depending solely on the scanning beam.

3.3.8.4 X-Ray Lithography

X-ray lithography (XRL) (0.8–0.1 nm) enables proximity masking to avoid the use of lenses. But it is restricted by two factors: (i) the lack of x-ray sources of sufficiently high intensity and (ii) the exposure of resists by secondary electrons produced by the high-energy photons.

3.3.8.5 Dip-Pen Nanolithography

A scanning probe nanopatterning technique, DPN is becoming a work-horse tool for researchers interested in fabricating and studying soft- and
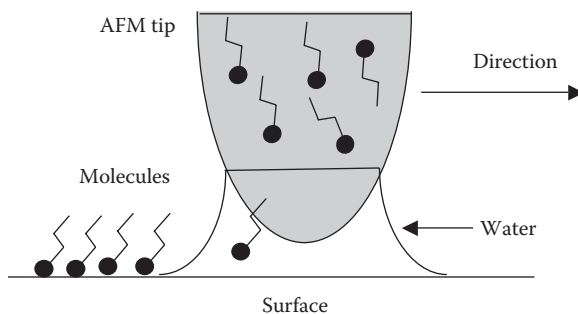
FIGURE 3.11
Principle of DIP-PEN nanolithography. (After Khanna, V. K., *Proceedings of IMS-2007, Trends in VLSI and Embedded Systems*, Punjab Engineering College, Chandigarh, 19, 2007.)

hard-matter on the sub-100 nm length scale. It is a direct-write technique that provides high-resolution patterning for a number of molecular and biomolecular inks on various substrates, for example, metals and semiconductors. Here, an AFM tip is used to deliver molecules to a surface via a solvent meniscus (curve in the upper surface of a standing body of liquid) (Figure 3.11). This meniscus naturally forms in the ambient atmosphere.

3.3.8.6 Nanoimprint Lithography

This method having the ability to pattern sub-25 nm patterns over large areas at high speed and low cost, comprises two steps: (i) *Imprint step:* A mold (a shaped cavity used to give a definite form to fluid material) with nanostructures on its surface is pressed into a thin resist layer coated on a substrate, duplicating the nanostructures on the mold in the resist film. During this step, the thermoplastic (a polymer that turns to a liquid when heated and freezes to a glassy state when cooled) resist is heated to a temperature above its glass transition temperature (the critical temperature at which the material changes its behavior from being "glassy" to being "rubbery") so that it can flow and can be readily deformed into the shape of the mold. (ii) *Pattern transfer step:* An anisotropic (directionally dependent) etching process, such as reactive ion etching (RIE, see following text), is used to remove the residual resist in the compressed area, transferring the thickness contrast pattern into the entire resist.

3.3.8.7 Nanosphere Lithography

Localized surface plasmon resonance (LSPR) and surface-enhanced Raman scattering (SERS) are optical nanosensing techniques. Nanoparticle size, shape, interparticle spacing, nanoparticle–substrate interaction, solvent, dielectric overlayers, and molecular adsorbates overwhelmingly affect the LSPR spectrum of nanoparticles. Also SERS technique is sensitive to the

working surface. This surface is often prepared with a distribution of metal nanoparticles. The shape and the size of such particles assertively affect the factor of enhancements of Raman scattering because they change absorption and scattering rates on the surface.

Photolithography has not been widely applied to nanostructure fabrication in view of its diffraction-limited resolution. Nanostructure fabrication processes that have hitherto been used are EBL and XRL. EBL is able to achieve a feature size as small as 1–2 nm, but is not adaptable for mass-volume fabrication because it is not a parallel process. On the other hand, XRL has achieved success in producing a comparatively higher volume but cannot strive for supremacy looking at the resolution limits of EBL. Nanosphere lithography (NSL) (Haynes and Van Duyne 2001) is an inexpensive, simple to implement, inherently parallel, high-throughput, nanostructure fabrication process, which systematically produces a two-dimensional (2D) array of periodic structures. It is a powerful fabrication technique to produce nanoparticle arrays with controlled shape, size, and interparticle spacing using self-assembled polystyrene $[(C_8H_8)_n]$ nanospheres as templates to replace photoresist mask for fabrication of nanostructures on various substrates. The NSL masks are created by spin-coating polystyrene nanospheres on the substrate of interest. After deposition of relevant material, that is, Ag or Au, the polystyrene nanospheres are removed from the substrate by dissolving them in CH_2Cl_2 with the aid of sonication.

3.3.9 Chemical Vapor Deposition

CVD has already been described earlier in the context of carbon nanotubes. It is a very important process in silicon technology for depositing layers like polysilicon, silicon dioxide (SiO_2), silicon nitride (Si_3N_4), etc., and will be discussed here in that backdrop. CVD is a generic name for a group of processes that use a chamber of reactive gas for depositing a solid material from a gaseous phase through a chemical reaction of vapor-phase precursors. Precursor gases (often diluted in carrier gases) are delivered into the reaction chamber at approximately ambient temperatures. As they pass over or come into contact with a heated substrate, they undergo reaction or decomposition forming a solid phase, which is deposited on to the substrate. The substrate temperature is critical and influences the reactions taking place.

What processes of semiconductor technology fall under the umbrella of CVD? CVD covers processes such as (i) atmospheric pressure CVD (APCVD); (ii) low-pressure or ultrahigh-vacuum CVD (LPCVD or UHVCVD)—LPCVD: subatmospheric pressures, 10–100 Pa, UHVCVD: 10^{-6} Pa; (iii) metal-organic CVD (MOCVD); (iv) plasma-enhanced CVD (PECVD); and (v) atomic layer deposition (ALD).

APCVD is performed in a reactor at temperatures up to ~400°C. As it is carried out at atmospheric pressure, it generally results in inferior film quality and conformality (the retention of angular relationships at each point)
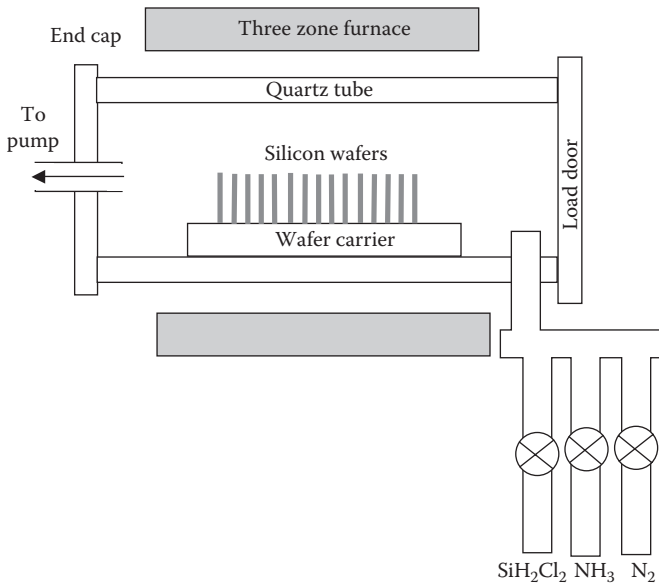
FIGURE 3.12
Low-pressure chemical vapor deposition (LPCVD) furnace for silicon nitride deposition on silicon wafers using dichlorosilane and ammonia.

of coating. The LPCVD process (Figure 3.12) produces layers with excellent uniformity of thickness and material characteristics. The main problems with the process are the high deposition temperature (higher than 600°C) and the relatively slow deposition rate. The PECVD process operates at lower temperatures (down to 300°C or less) due to the extra energy supplied to the gas molecules by the plasma in the reactor. However, the quality of the deposited films tends to be inferior to processes running at higher temperatures. In MOCVD, atoms required in the crystal are combined with complex organic gas molecules and passed over a hot semiconductor wafer. The heat breaks up the molecules and deposits the desired atoms on the surface, layer by layer. By varying the composition of the gas, the properties of the crystal are manipulated at an almost atomic scale. The reaction for GaN growth is

$$Ga(CH_3)_3 + NH_3 \rightarrow GaN + CH_4 \qquad (3.34)$$
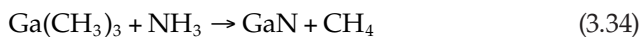
In what way is ALD a modification of CVD? ALD, sometimes called atomic layer epitaxy (ALE), pulsed CVD, or atomic layer chemical vapor deposition (ALCVD), is a modified version of the CVD process having the following salient features: (i) gaseous precursors are introduced sequentially to the substrate surface (Jones and Hitchman 2009). (ii) Furthermore, the reactor is

purged with an inert gas, or evacuated, which removes any excess precursor molecules and volatile (evaporating readily at normal temperatures and pressures) by-products from the reaction chamber, thus eliminating the possibility of occurrence of unwanted gas phase reactions.

How does ALD differ from conventional CVD? In marked contrast to traditional thermal CVD, which involves pyrolysis (thermolysis: thermochemical decomposition of materials at elevated temperatures in the absence of oxygen) of precursor molecules, (i) ALD proceeds through surface exchange reactions, such as hydrolysis (subjecting to chemical action of water), between chemisorbed metal-containing precursor fragments and adsorbed nucleophilic reactant molecules; a *nucleophile* is a chemical compound or group that is attracted to nuclei and tends to donate or share electrons; (ii) it must be noted that the chemical reactions leading to film deposition in ALD take place exclusively on the substrate at temperatures below the thermal decomposition temperature of the metal-containing precursor; and (iii) also, the gas-phase reactions are not important.

By CVD, various films required in CMOS process are deposited, notably polysilicon and different dielectric films.

Polysilicon deposition reaction

$$SiH_4 \rightarrow Si + 2H_2 \, (575°C - 650°C) \tag{3.35}$$
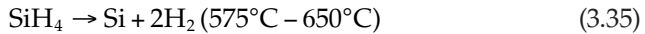
Silicon dioxide deposition reactions

$$SiH_4 + O_2 \rightarrow SiO_2 + 2H_2 (< 500°C); \quad Si(OC_2H_5)_4 \rightarrow SiO_2$$

$$+ \text{by-products} (650°C - 750°C); \quad SiCl_2H_2 + 2N_2O \rightarrow SiO_2 + 2N_2 + 2HCl \tag{3.36}$$
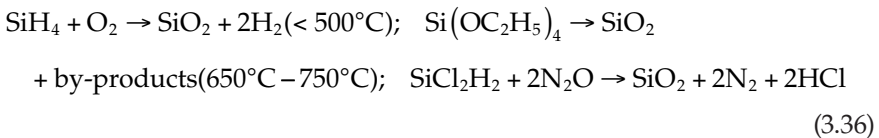
SiH_4 is silane. $Si(OC_2H_5)_4$ is tetraethoxysilane (TEOS). $SiCl_2H_2$ is dichlorosilane.

Silicon nitride deposition reactions

$$3SiH_4 + 4NH_3 \rightarrow Si_3N_4 + 12H_2 (700°C - 900°C);$$

$$3SiCl_2H_2 + 4NH_3 \rightarrow Si_3N_4 + 6HCl + 6H_2 (700°C - 800°C) \tag{3.37}$$
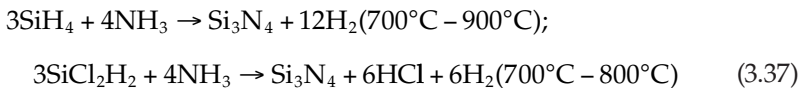
What about the higher dielectric constant (high-κ) film deposition? The silicon oxynitride is the starting step toward the use of a material with high-κ and possessing satisfactory insulating properties as the gate insulator. It enables reduction of gate tunneling current using materials compatible with silicon technology: combining SiO_2 and silicon nitride (Si_3N_4) in the insulating layer. The results achieved depend on the processing of the mixture. Various mixtures of the

form SiN_xO_y, where $3x + 2y = 4$, known as *oxynitrides*, are used as the gate insulator. A gate–substrate capacitance equivalent to an oxide insulator about 2 nm thinner is obtained for a given tunneling current density. Another method superimposes a layer of silicon nitride above a thin SiO_2 layer as the gate insulator, thereby preserving the desirable qualities of the contact of SiO_2 on silicon. The nitride has a dielectric constant of 7 as compared with 3.9 for the oxide, and the use of silicon nitride is well established in silicon technology.

What are the other dielectrics available? As already stated, other dielectrics include hafnium and zirconium silicates and oxides that are typically deposited using ALCVD. The dielectric constant of HfO_2 and SrO_2 is ~20–25, which is a factor of 5–6 times that of silicon dioxide. Rare earth (RE) oxides such as crystalline praseodymium oxide (Pr_2O_3) deposited by MBE have been successfully introduced into a CMOS process. RE elements are a collection of 17 chemical elements in the periodic table, specifically the 15 lanthanoids plus scandium and yttrium. From capacitance–voltage measurements, a dielectric constant of $\kappa = 36$ has been calculated.

3.3.10 Wet Chemical Etching and Common Etchants

This is the simplest etching technology using a container with a liquid solution that will dissolve the material in question. Buffer HF is commonly used for etching silicon dioxide selectively while orthophosphoric acid (H_3PO_4)-based compositions are employed for aluminum etching. Etchant for gold is potassium iodide (KI) in iodine (I_2). Silicon etchants are mixtures of HF, HNO_3, and acetic acid (CH_3COOH).

3.3.11 Reactive Ion Etching

Can etching be done without liquid etchants? Yes, RIE is a process that uses dry chemical and physical processes to etch away desired materials. *What is the essence of RIE?* In RIE, the substrate is placed inside a reactor in which a mixture of gases is introduced (Figure 3.13). A plasma (a gas of positive ions and free electrons containing approximately equal positive and negative charges) is struck in the gas mixture using an RF (radio frequency) power source, breaking the gas molecules into ions. The ions are accelerated toward and react at the surface of the material being etched, forming another gaseous material. This is known as the *chemical part of RIE*. There is also a *physical part* that is similar in nature to the sputtering deposition process. If the ions have high enough energy, they can knock atoms out of the material to be etched without a chemical reaction. This dislodging of atoms by impact or collision constitutes the physical part of the etching. *Synergism of chemical reaction and ion bombardment leads to high etch rates*.

How does RIE compare with wet etching? Compared to wet etching, RIE has higher anisotropy (having dissimilar properties in different directions), better uniformity and control, and better etch selectivity.
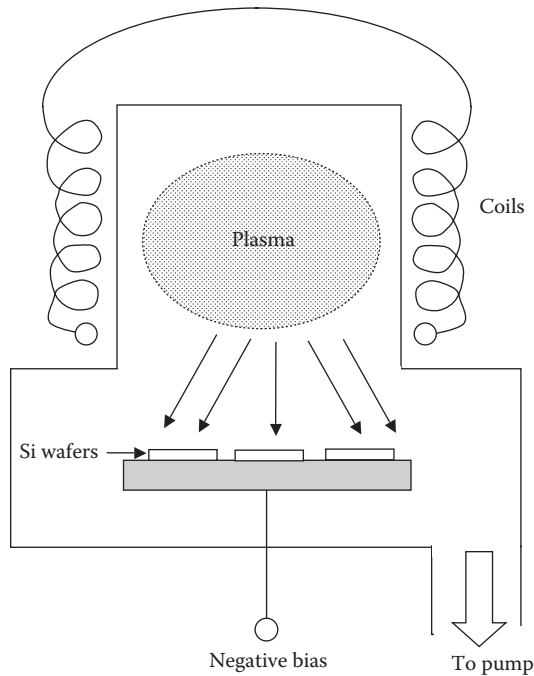
FIGURE 3.13
Transformer-coupled plasma RIE system.

3.3.12 Focused Ion Beam Etching and Deposition

Focused ion beam (FIB) technology is a widely used tool for microelectronics. An FIB system using liquid metal ion sources is capable of forming very small probes with high current densities. The ions strike the specimen to remove materials through a physical sputtering process. A combination of a FIB and a reactive gas also causes various physical effects on the specimen. Gas molecules above a specimen may be dissociated resulting in a local etching or deposition of the specimen.

FIB systems operate in a similar fashion to a scanning electron microscope (SEM); the exception is that instead of a beam of electrons, FIB systems use a sharply focused beam of gallium (Ga^+) ions that is operated at low beam currents for imaging or high beam currents for site-specific sputtering or milling. The gallium (Ga^+) primary ion beam strikes the sample surface and sputters a small amount of material, which departs from the surface as either secondary ions or neutral atoms. The primary beam also produces secondary electrons. As the primary beam rasters on the sample surface, the signal from the sputtered ions or secondary electrons is collected, forming an image.

By spraying compound gas on the sample surface near the ion beam irradiation area, deposition is performed locally. Secondary electrons are

generated when primary ions are irradiated. Secondary ions contribute to the decomposition of compound gas, which splits up into gaseous and solid components. The gaseous component is evacuated in vacuum, but the solid component piles up on the sample surface. From this build up, maskless deposition is performed selectively on ion beam irradiation areas.

While electron beam deposition can regulate the damage to the sample because there is no sputtering effect with beam irradiation, the formation speed of the deposition film, or the deposition rate, appears to be low.

By increasing the ion beam amount that would increase the sputtered atom amount, etching is performed on the sample surface. Maskless processing, which selectively etches away those parts where ion beam is irradiated, is enabled. Using this technique, etching is performed on predetermined spots of the sample.

3.3.13 Metallization

Metallization is the process that provides contacting regions for separate devices as well as interconnects individual devices together by means of metallic lines (microscopic wires) to form circuits. *What are the principal requirements of metal film?* Desirable features of metal interconnections are (i) high adhesion to substrate, (ii) low ohmic and contact resistances (the contributions to the total resistance of a material that comes from the electrical leads and connections), and (iii) reliable long-term operation. *Which metal is most popular for metallization?* Aluminum is a popular metal used to interconnect ICs. *How is aluminum deposited?* Al metal layers are usually deposited through PVD by sputtering. In a magnetron (a high-powered vacuum tube that generates microwaves using the interaction of a stream of electrons with a magnetic field) sputtering application (Figure 3.14), the high voltage is delivered across a low-pressure gas (usually argon) to create high-energy plasma. This plasma emits a colorful halo of light often referred to as a "glow discharge" and consists of electrons and gas ions. These energized plasma ions strike a target composed of the desired coating material. The force causes atoms to eject from the target material, strike and bond with those of the substrate. Because sputtering takes place in a high-energy environment, it creates a virtually unbreakable bond between the film and its substrate at the atomic level, creating one of the thinnest, adherent, most uniform, and cost-effective films possible. Another method used for metallization is electron beam (E-beam) evaporation in which a beam of electrons emitted by a tungsten filament under high vacuum is focused on the source material.

The replacement of aluminum (resistivity $= 2.65 \times 10^{-6}$ Ω cm), long used for wiring of integrated circuits, with lower resistivity copper (1.67×10^{-6} Ω cm) to reduce RC delays (resistive–capacitive delays) has been successful. Copper is also much less susceptible to electromigration, the movement of atoms carrying large electrical currents, than aluminum.
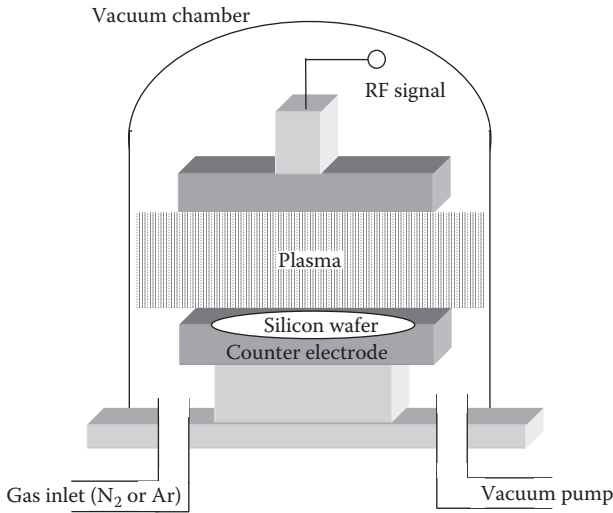
FIGURE 3.14
RF sputtering system.

Are copper interconnections made in the same way as aluminum? No, the process of manufacturing of copper (Cu) interconnections is not the same as aluminum-based interconnections because of possible oxide poisoning and diffusion problems. (i) A barrier layer is needed to prevent the diffusion of copper into neighboring dielectric layers and the silicon substrate. (ii) A cladding around on-chip copper interconnects is formed. It was found necessary to have a lining of tungsten (W) around the copper to prevent reaction with the SiO_2.

Electroplating has been used to create copper interconnects due to its superior trench filling capability, higher reliability, and low cost. The IBM process provides void-free copper electrofilling in extremely narrow geometry, high aspect ratio IC structures.

3.3.14 Dicing, Wire Bonding, and Encapsulation

During dicing, the wafer saw, consisting of a blade embedded with diamond particles that rotates at a very high speed, passes through the wafer at boundaries between dies known as *saw streets*, established during wafer fabrication. The dicing machine is programmed to drive the saw blade through the saw streets at a defined spindle speed, saw rate, and depth, separating the wafer into individual dies. The blade must be carefully aligned with saw streets to avoid any chip loss at this stage where the wafer has undergone all the processes.

Wire bonding is the process of providing electrical connection between the silicon chip and the external leads of the semiconductor device by attaching

a very thin wire, usually 25–75 µm in diameter, from one connection pad to another, completing the electrical connection in the electronic device. The most frequently used method of joining the wires is *ultrasonic welding*, a high-frequency ultrasonic acoustic vibration is locally applied to workpieces being held together. *Thermocompression wire bonding* requires the application of a high force on the surface along with a high temperature, around 300°C. It provides excellent, reliable Al–Au bonds with flexibility in the bonding direction allowed. The wire material is Au but the pad can be either Au or aluminum. Depending on their shapes, there are two basic types of wire bonds: ball/wedge and wedge/wedge. About 90% of all assemblies are produced using ball bonds and about 10% are done with wedge bonds.

Package is a case or molding forming the body of a commercial semiconductor device. *Integrated circuit encapsulation* refers to the design and manufacturing of protective packages for these circuits. IC packaging materials include plastics, ceramics, laminates (materials that are constructed by uniting two or more layers of material together), metal, etc. IC package categories are as follows: (i) pin-through-hole (PTH), where pins are inserted through holes in the circuit board; and (ii) surface-mount-technology (SMT), in which packages have leads that are directly soldered to the exposed metal lands on the surface of the circuit board.

Reliable operation of nanosensors in harsh environments such as elevated temperatures, high pressures, and aggressive chemical media poses stringent requirements on packaging, which cannot be met using polymer-based technologies. Ceramic technologies, especially low-temperature co-fired ceramics (LTCC), an evolution of thick film technology where the multilayer substrate is co-fired with the other layers (resistors, conductors, etc.), offer a trustworthy platform to build stable and reliable packages. LTCC allows fabrication of complicated 3D structures with fluidic channels for liquids and/or gases, while simultaneously permitting 3D electrical circuits in the same device.

3.3.15 IC Downscaling: Special Technologies and Processes

3.3.15.1 Downscaling Trends

During the past several decades, CMOS (complementary metal-oxide semiconductor) integrated circuits have been at the frontiers of research and development globally. Research activities in this field have expanded exponentially and a revolution is apparent leading to reliable, low-cost ICs for consumer electronics, computer, and telecommunication sectors (Khanna 2004, 2007a). Deep submicron (DSM) technology involves the use of <0.35 µm (350 nm) and ultra-deep submicron (UDSM) <0.25 µm (250 nm) feature sizes in ICs. State-of-the-art nano CMOS circuits follow 90 nm design rules today, and they are likely to reach 32 nm by the beginning of the next decade. With

reference to the *International Technology Roadmap for Semiconductors* (1999 edition), a minimum channel length of 20–22 nm for the MOS transistor is predicted for the year 2014. Although the resolution of optical lithography is gradually increasing, there is no known and proven method for structure sizes below 100 nm. EBL or XRL can meet this goal but low-cost solutions for mass production are still elusive.

3.3.15.2 SOI-MOSFETs

The monumental growth of silicon technology is supported on the foundation of new fabrication technologies and materials. Silicon-on-insulator metal-oxide-semiconductor field-effect transistors (SOI-MOSFETs) (Figure 3.15) offer advantages such as easy device isolation, absence of latch-up (a failure mechanism in the form of a particular type of short circuit that occurs in an improperly designed device or circuit), extremely low junction capacitance, improved transconductance, better subthreshold slope, suppression of the short channel effect (an effect whereby a MOSFET, in which the channel length is the same order of magnitude as the depletion-layer widths of the source and drain junctions, behaves differently from other MOSFETs), and high temperature operation capability such as 350°C. The isolation advantage makes the SOI process sequence free from trench or well formation steps. Two methods of making wafers for the SOI devices commonly used in practice are *SIMOX* (separation by the implantation of oxygen) and *smart cut* processes.
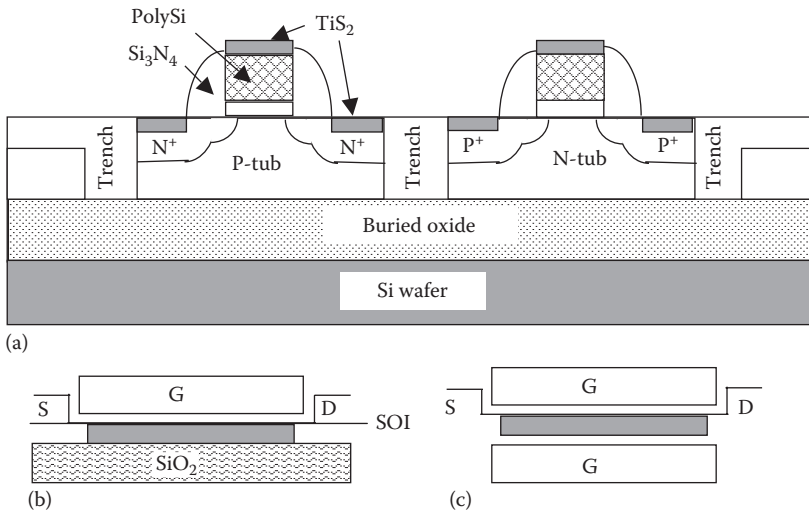
FIGURE 3.15
(a) Trench-isolated SOI CMOS structure, (b) single-gate FET, and (c) double-gate FET. (After Khanna, V. K., *Proceedings of IMS-2007, Trends in VLSI and Embedded Systems*, Punjab Engineering College, Chandigarh, 19, 2007.)

3.3.15.3 SIMOX Process

The SIMOX method employs oxygen ion implantation to produce a high concentration of oxygen in a layer beneath the silicon surface. High temperature annealing of the implanted wafer causes crystallization of a silicon layer at the surface that is separated from the bulk by a layer of SiO_2. The buried insulator is called *buried oxide* (BOX).

3.3.15.4 Smart Cut Process

In the Smart Cut Process, a layer of SiO_2 is formed on a silicon wafer A. Proton implantation through the oxide layer formed into silicon produces a damaged layer at the end of the penetration range of protons. The oxidized surface is bonded to another silicon wafer B. The bonded wafer pair is heated so that the two wafers split along the hydrogen-implanted plane; this procedure is termed *hydrogen slicing*. Wafer B becomes the substrate of the SOI wafer and the remaining wafer A is reused as a substrate for another SOI wafer. Wafer B is annealed and polished to obtain the required SOI wafer for device fabrication.

3.3.15.5 Strained Silicon Process

Stresses influence the bandgap and carrier mobility (drift velocity per unit applied electric field) in silicon. The pursuit of high mobility has led to the introduction of germanium into silicon. Silicon and germanium are miscible elements. They form alloys with a range lattice parameter that increases from silicon to germanium. Silicon is grown epitaxially (e.g., by using metal-organic vapor phase epitaxy, MOVPE) on a Ge–Si alloy with a larger lattice constant (the constant distance between unit cells in a crystal lattice) and strained in such a way that current in the plane of the channel is carried by high-mobility, low-mass electrons. In particular, charge carriers in germanium have a smaller effective mass and higher mobility than those in silicon. Thus, silicon technology has returned to a thin layer of germanium as a means to obtain higher mobility.

The *IBM strained silicon process*, called *dual stress liner* enhances the performance of both types of semiconductor transistors, called N-channel and P-channel transistors, by stretching silicon atoms in one transistor and compressing them in the other. *Strained silicon* is a layer of Si in which the Si atoms have been stretched beyond their normal interatomic distance. Because of the natural tendency for atoms inside compounds to align with one another, the atoms in the silicon layer align with those in the SiGe layer where the atoms are farther apart. Hence, the links between the silicon atoms become extended, thereby leading to strained silicon.

In strained silicon, the interatomic spacing in the plane of the wafer is greater than in unstrained Si. Biaxial (having two axes) distortion of the Si lattice alters its electronic band structure. Hence, electron and hole mobilities

increase. Electrons move at 70% faster speeds so that strained silicon devices can switch 35% faster.

It may be noted that the introduction of germanium in the critical areas of the integrated circuits provides an alternative means of improving chip performance from the traditional method of simply shrinking circuitry. This is becoming increasingly important as further miniaturization becomes more difficult and yields diminishing returns. Future silicon technology is likely to depend on locally strained silicon channels.

3.3.15.6 Top–Down and Bottom–Up Approaches

The *top–down route* of the silicon technology has been running easily until its basic step of optical lithography has met its physical limits (minimum feature size around the light wavelength). Quest for new technological options has begun. Nanotechnology has provided the answer by developing techniques to handle individual molecules. In opposition to the present downscaling of CMOS ICs, it is necessary to reverse the process flow to bottom–up approach.

Bottom–up approach is nonlithographic in contrast to the top–down approach using expensive photolithographic, thin-film deposition, etching, and metallization processes. Its main tools are chemical synthesis and self-assembly. *Inexpensive chemistry* is used to promote self-assembly of complex mesoscopic architectures; the *mesoscopic* world lies in between the microscopic and the macroscopic world. Further, in molecular components made by organic synthesis, smaller variations are observed in structural parameters and chemical composition. Molecular species are tailored for different device applications through *surface molecular engineering*.

Molecular electronics is the stage of electronics in which the data-handling device will be coincident with the constituting material (the molecule). Instead of imagining about the ideal situation of total molecular electronics, the hybrid molecular electronics approach, in which microelectronic circuits host molecular devices grafted to the silicon surface, seems the most realistic one for future electronics. Thus, top–down and bottom–up approaches will not be antagonistic but complementary.

3.3.15.7 DNA Electronics

What striking qualities of DNA (deoxyribonucleic acid) molecules favor their application in electronics? (i) Organic molecules like DNA can be used either as active components through which electron current flows such as active channel of an FET or quantum device; or as passive devices, for example, charge storage medium of a memory. (ii) DNA possesses complementarity. This is a *self-assembling ability*. DNA forms a self-assembled structure that always has hydrogen-bonded base pairs of guanine ($C_5H_5N_5O$) and cytosine ($C_4H_5N_3O$), and of adenine ($C_5H_5N_5$) and thymine ($C_5H_6N_2O_2$), the four bases found in DNA and represented by the letters G–C–A–T. This structure can

be controlled by programming the base sequence. Due to the self-assembling property of DNA, it can be highly integrated error-free without using microfabrication technologies. (iii) In addition, DNA has an address with 0.34–0.36 nm intervals. It serves as an information material that can be used as a template for aligning molecules and clusters in nanoscale without any ultrafine processing. (iv) DNA is an insulator, and periodic DNA can be a wide-bandgap semiconductor. (v) Since the electric conduction in DNA is controlled by doping, doped DNA can be regarded as a nanoscale semiconducting molecular wire. (vi) In the atmosphere, the electric conduction of DNA is dominated by ion conduction in the water layer adsorbed onto it.

Thus, DNA serves as a nanoscale electronics material that works as an electronic transport material in 1D molecular wires if used in vacuum and an ion transport material if used in the atmosphere.

3.3.15.8 Spintronics

Spin-based electronics, often called *spin-electronics* or *spintronics* employs the spin degrees of freedom in solid-state systems (Tanaka 2005). Spintronics harnesses an electron's spin (quantum property with either an "up" or a "down" orientation) for data encoding and processing. *What types of controlling fields are used in electronics and spintronics?* While electronics uses electrical fields to push electrons for reading information in a current, spintronics uses *magnetic fields* to transport electrons. Spintronics could lead to the development of computers using less power and with more data storage capability.

The spin MOSFET consists of a MOS gate with a Si channel (Figure 3.16). The source and drain materials are half metallic ferromagnets with 100% spin polarization at the Fermi energy, implying that carriers are completely spin polarized. The source/channel and drain/channel contacts are Schottky barriers (potential barriers formed at a metal–semiconductor junction that have rectifying characteristics). They allow efficient spin injection and detection and also act as blocking contacts for stopping the drain–source current in the off-state of FET. Spin injection utilizes the strong, short-range quantum mechanical exchange interaction of the injected spin-polarized
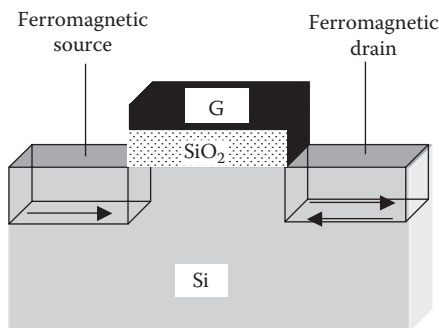
FIGURE 3.16
Spin MOSFET.

electrons with the atomic spins on the atoms. From the half-metallic source, highly spin-polarized carriers are injected by tunneling across the Schottky barrier into the Si channel. The half-metallic drain selectively extracts the spin-polarized carriers from the channel. This happens only when the spin configuration between the ferromagnetic source and drain is parallel. Hence, the output current depends on the relative magnetization configuration of the source and drain. Since the Si channel can be intrinsic, P-type, N-type, or both P-type and N-type, spin MOSFETs, and, therefore, CMOS logic gates can be designed. These are compatible with the CMOS technology and promising for computing applications.

3.4 MEMS and NEMS Division

3.4.1 Surface and Bulk Micromachining

Surface and bulk micromachining processes are used to create microstructures in MEMS devices (Figure 3.17) (Khanna 2007b). *How do these processes differ?* The difference between surface and bulk micromachining is that instead of etching the silicon substrate (as done in bulk micromachining), surface
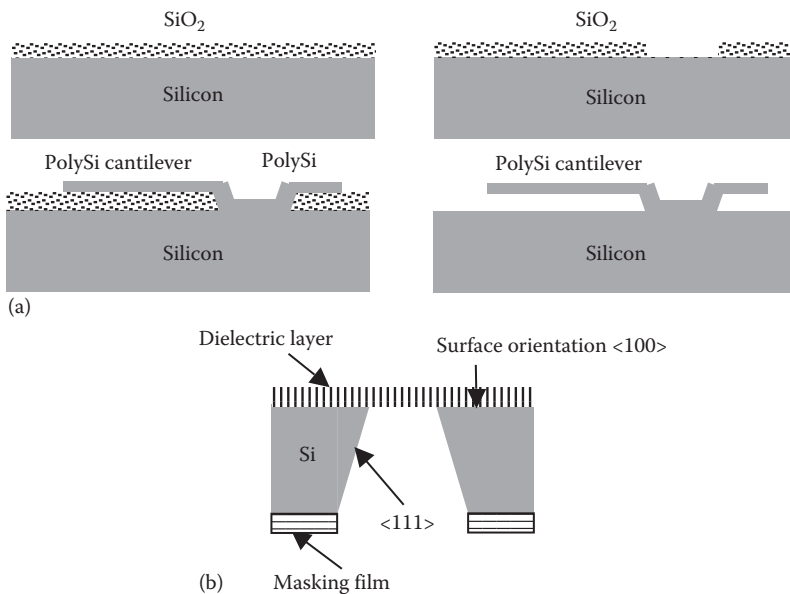
FIGURE 3.17
(a) Surface micromachining for cantilever fabrication and (b) bulk micromachining. (After Khanna, V. K., *Proceedings of IMS-2007, Trends in VLSI and Embedded Systems*, Punjab Engineering College, Chandigarh, 317, 2007.)

micromachining etches away layers deposited on top of the silicon substrate. Surface micromachining combines structural materials and sacrificial layers for making MEMS structures on the silicon surface. The process starts with a silicon wafer, upon which structural and sacrificial layers are deposited. *What are structural and sacrificial layers? Structural layers* are the layers that form the desired structures. *Sacrificial layers* are the layers that are etched away, and are used to support the structural layers until they are etched away. Typically, a sacrificial layer of silicon dioxide is formed by a combination of thermal and CVD processes. Phosphosilicate glass (PSG) is also often used as a sacrificial layer because of its high etch rate in hydrofluoric acid (HF). After the polysilicon structural layers are selectively deposited on top of the sacrificial layer, the silicon dioxide is etched away using HF. This process in which polysilicon is the commonly used structural material while SiO_2 and related glasses like PSG act as the sacrificial material that is etched away at a later stage of the process is useful for creating cantilever beams, bridges, and sealed cavities.

Bulk micromachining selectively etches the silicon substrate to create microstructures in MEMS devices. Typically, a layer of silicon dioxide is patterned onto a silicon wafer using a mask. The silicon dioxide is patterned to protect certain areas of the silicon substrate from etching. Thus, in bulk micromachining, 3D structures are etched into the bulk of silicon in contrast to building the features layer by layer on the surface of a silicon wafer.

3.4.2 Machining by Wet and Dry Etching Techniques

Etching is a process in which material is removed from selected regions of the substrate (Khanna 2007c). In *wet etching*, liquid etchants are used to remove material. Wafers are immersed in a tank containing the reactants. Wet chemical etching is of two types: isotropic and anisotropic. *Isotropic etchants* attack Si at the same rate in all directions. *Anisotropic etchants* attack the silicon wafer at dissimilar rates along different directions. *Which etching has more control on the resulting shapes?* Anisotropic etching is direction-sensitive and so there is more control of the shapes produced. An etching process that progresses in one direction, for example, vertical only, is termed "completely anisotropic." *What etchants are commonly used for silicon?* The common Si etchants are inorganic aqueous solutions of KOH, NaOH, RbOH, CsOH, NH_4OH, hydrazine $((N_2H_4)_n)$; and organic etchants like EDP (Ethylenediamene pyrocatecol) consisting of ethylenediamine $(C_2H_4(NH_2)_2)$, pyrocatechol $(C_6H_4(OH)_2)$, pyrazine $(C_4H_4N_2)$, and water; and TMAH (Tetramethyl ammonium hydroxide), a quaternary ammonium salt with the molecular formula $(CH_3)_4NOH$. *Which amongst these are the most popular etchants, and what are their typical etching rates?* Amongst these, the most commonly used etchants include KOH, TMAH, and EDP. Fifty grams of KOH in 100 mL H_2O/isopropanol (C_3H_8O) at 50°C gives an etch rate ~ 1 μm min^{-1}. Biswas and Kal (2006) reported that 22 wt% KOH solution provides maximum silicon etch rates of 89.2 and 88.1 μm h^{-1} for N-type and P-type silicon, respectively, at 80°C. Biswas et al. (2006) also found that

TABLE 3.1

Different Silicon Etchants for Bulk Micromachining

Etchant	Features and Advantages	Disadvantages
Potassium hydroxide (KOH), an alkali metal hydroxide	Nontoxic, economical, requires simple etching setup, provides high silicon etch rate, high degree of anisotropy, moderate Si/SiO_2 etch rate ratio and low etched surface roughness	Damages exposed aluminum metal lines very quickly. Not CMOS compatible due to the presence of K^+ alkali metal ions in it
Ethylenediamene pyrocatecol (EDP), a diamine-based silicon etchant	Moderate silicon etch rate, high Si/SiO_2 etch rate ratio, low degree of anisotropy and partly CMOS compatible	Toxic, requires careful handling, special safety measures and therefore a complex etching apparatus. Ages quickly
Tetramethyl ammonium hydroxide (TMAH), a quaternary ammonium hydroxide based silicon etchant	A nontoxic, CMOS compatible organic solution, moderately high silicon etch rate and high selectivity to masking layers	High cost and requires a complex etching setup. Gives rough etched silicon surfaces. But when doped with suitable amounts of silicic acid and AP, $[(NH_4)_2S_2O_8)]$, provides full aluminum passivation along with smoothly etched surfaces

undoped 5% TMAH yields etch rates of 10–60 μm h^{-1} and surface roughness between 3 and 1.5 μm with etching temperature changing from 50°C to 80°C. Five percent TMAH solution doped with 38 gm L^{-1} silicic acid (H_4O_4Si) and 7 gm L^{-1} AP [ammonium peroxodisulfate: $(NH_4)_2S_2O_8$)] gives a high <100> Si etch rate of 70 μm h^{-1}.

In wet etching, the choice of a particular silicon etchant for fabricating structures depends not only on the etch rate selectivity and anisotropy but also on its compatibility with integrated circuit technology. TMAH is fast gaining popularity in MEMS as an alternative to KOH and EDP. A comparison is made in Table 3.1.

How is micromachining done by dry etching and what are the common types of dry etching? As briefly described before, dry etching is done by a plasma without using any liquids. Three types of dry etching, namely, reactive ion etching, sputter etching, and vapor phase etching, are distinguished in Table 3.2.

3.4.3 Deep Reactive Ion Etching

What is DRIE? A special subcategory of RIE that is rapidly gaining reputation is *deep RIE* (*DRIE*), which has virtually changed the MEMS scenario. In this process, etch depths of hundreds of microns are achieved; at the same time, almost vertical sidewalls are ensured.

What is the remarkable feature of DRIE? It prevents lateral etching of the silicon resulting in highly anisotropic etch profiles at fast etch rates and with large

TABLE 3.2

Types of Dry Etching

Sputter Etching (Physical Etching)	Vapor Phase Etching	Reactive Ion Etching
Material is removed by bombarding the Si substrate with gas ions. Etchant is an inert gas like argon giving a slow etch rate, typically tens of nanometer per minute	Si wafer to be etched is placed inside a chamber in which one or more gases are introduced, and the material is dissolved at the surface in a chemical reaction with the gas molecules. Silicon etching using xenon difluoride (XeF_2), BrF_3, and ClF_3 is isotropic in nature. Silicon reacts with the gas forming SiF_4. Etch rate in XeF_2 is ~0.27 micron min^{-1}	By striking a plasma in the gas mixture (SF_6) using an RF power source, the gas molecules are ionized, which are accelerated toward, and react at the surface of the material being etched, forming another gaseous material. This is the chemical part of RIE. It also has a physical part. If the ions have high energy, they knock atoms out of the material to be etched without a chemical reaction. Since the chemical part is isotropic and the physical part highly anisotropic, the combination forms sidewalls that have shapes from rounded to vertical. Etch rate is ~2–3 micron min^{-1}

aspect ratio compared to wet chemical etching. The *aspect ratio* of a given feature, like a trench, is defined as the ratio between the depth of the trench and the width.

What is the sequence of steps in a DRIE process? In the time multiplexing (the process of dividing up one time slot into smaller time slots) dry etching scheme, the etching and passivating gases are made to flow independently. An etching cycle and a passivating cycle are alternately applied to the machine. During the etching cycle of duration ≤12 s, a shallow trench is formed in the silicon substrate with an isotropic profile of fluorine (F)-rich glow discharge. In the ensuing passivation cycle of duration ≤10 s, a protective fluorocarbon (organo-fluorine compound that contains only carbon and fluorine bonded together in strong carbon–fluorine bonds) film is formed on all the surfaces. In the subsequent etching cycle, ion bombardment removes the passivant from all the horizontal surfaces continuing the etching process but fails to do so on the vertical sidewalls so that the walls are not etched allowing the etching profile to evolve in a highly anisotropic fashion. *What is this alternation of cycles known as?* This alternation of etching and passivating cycles is known as the *Bosch process* after the German company that invented and patented it (Figure 3.18).

Is there any other version of DRIE process? Yes, another variant of DRIE uses cryogenic (very low temperature below –150°C) process. Distinction between the two DRIE processes is clarified in Table 3.3.

In brief, dry processes such as RIE have been widely used for making narrow vertical-wall structures. The addition of C_xF_y polymerizing species to a
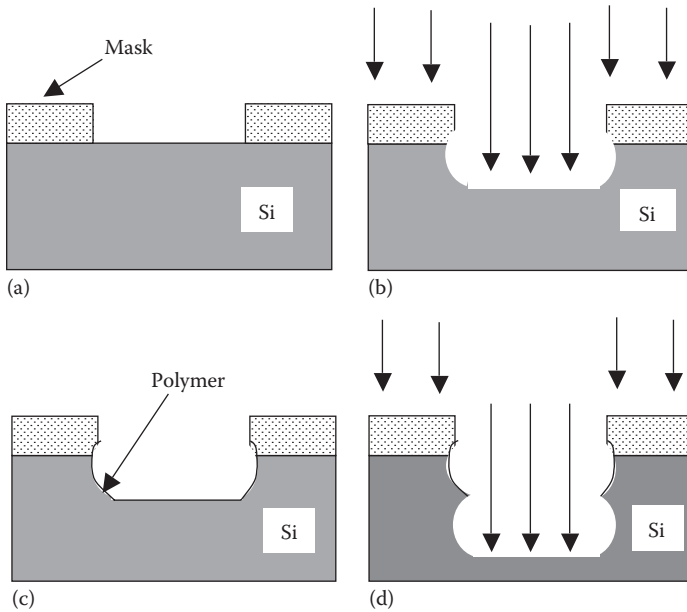
FIGURE 3.18
Bosch process sequence: (a) photolithography for etching window definition, (b) etching, (c) polymer deposition, and (d) etching to greater depth. (After Khanna, V. K., *Proceedings of the All India Conference on Recent Developments in Manufacturing & Quality Management (RDMQM - 2007)*, Punjab Engineering College, Chandigarh, 46, 2007.)

TABLE 3.3

DRIE Processes

Bosch Process	Cryogenic Process
"Bosch process," a patented process developed by Robert Bosch GmbH is based on alternating multiple steps of etching by high-density plasma (SF_6) with sidewall passivation by an etch-resistant polymer, usually a fluorocarbon with composition C_xF_y like C_4F_8 (Figure 3.18). Etching rates ~6–12 mm min^{-1}, which are several fold higher than wet etching	The mechanism is condensation of reactant gas ($SF_6 + O_2$) on the sidewalls of the reactor by cooling while condensation at the bottom is removed by ion bombardment. Advantage is absence of polymer contamination. Also provides lower sidewall roughness

fluorine-containing plasma is a flexible high-etching rate approach wherein the ratio of vertical to lateral etch rate is adjusted providing fine-tuning of the profile from vertical to tapered. But it must be emphasized that these processes require relatively high initial equipment and running costs.

What factors determine the suitability of an etching technique for a particular application? Appropriateness of a particular technique for a specific application is governed by several factors. Techno-economic aspects such as extent of

control over etching profile vis-à-vis equipment cost, installation, maintenance, and infrastructural expenses are prime considerations. Anisotropic wet etching of <100> silicon is an established technique that has been widely used for realization of microstructures. The use of dry etching is restricted by prohibitively high cost to specialized applications where other methods are inadequate. Dry etching is a high-precision and controlled process but also high-cost-enabling technology.

Under what circumstances, is dry etching essential? If feature resolution in thin film structures is of primary concern, or vertical sidewalls are needed for deep etchings in the substrate, dry etching must be irresistibly done. *Under what conditions can dry etching be dispensed with*? If the above are of secondary interest and price of the process and device must be reduced, the use of dry etching is minimized.

3.4.4 Front- and Back-Side Mask Alignment

The fabrication of MEMS/NEMS devices often involves double-side mask alignment. Approaches to implement this kind of alignment are as follows: (i) Conventional alignment of a pattern on the bottom surface of a wafer to one on the top surface, that is, front-to-back alignment, is done with an infrared mask aligner for aligning through the substrate. (ii) Another method for double-sided alignment requires the use of a fixture that holds two masks in alignment with each other. The substrate is placed between the two masks and imaged. (iii) Bottom objectives are used to align the mask features to the back of the wafer by aligning a captured image of the mask with the real image of the wafer.

3.4.5 Multiple Wafer Bonding and Glass–Silicon Bonding

In MEMS/NEMS fabrication, it is frequently necessary to bond two silicon wafers together or a silicon wafer with a glass wafer. For this purpose, various bonding approaches are as follows: (i) *Anodic bonding:* such bonding between silicon and glass is done at 450°C at 500–1000 V to attract Na$^+$ ions in glass to the negative electrode where they are neutralized (Figure 3.19). This forms a space charge at the glass–silicon interface producing a strong electrostatic attraction between Si and glass whereby oxygen ions move from glass to glass–silicon interface forming silicon dioxide (SiO_2) and hence a permanent bond. The typical bond strength achievable by this method is about 5.6 kg cm^{-2}. Thus, anodic bonding is a method of hermetically and permanently joining a glass substrate with a silicon wafer without using adhesives. (ii) *Silicon direct bonding or silicon fusion bonding*: it is based on initial bonding by hydroxyl groups present on silicon wafer surfaces prepared by RCA (Radio Corporation of America) cleaning before bonding. Mechanical spacers are placed at the edges. Removing the spacers allows a single bonding wave to propagate from the center of the wafers. The wafers are annealed above 1000°C so that the hydroxyl groups from water molecules create Si–O–Si bonds as
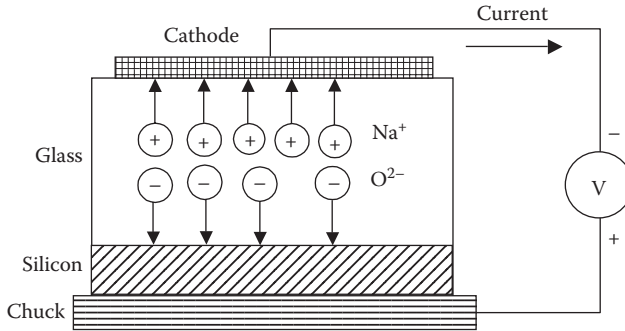
FIGURE 3.19
Glass–Si anodic bonding. (After Khanna, V. K., *Proceedings of IMS-2007, Trends in VLSI and Embedded Systems*, Punjab Engineering College, Chandigarh, 317, 2007.)

H_2 diffuses away. (iii) *Intermediate layer bonding*: it includes eutectic and glass-frit materials made from 100% crushed post-consumer recycled glass. Gold has a eutectic temperature of 363°C with silicon. Gold is evaporated or plated on one of the wafers. For eutectic bonding, the wafers are held slightly above eutectic temperature (the temperature in a two-component mixture where a liquid solution and both pure solids exist at a fixed pressure). In glass-frit bonding, a thin glass layer such as lead borate $[Pb(BO_2) \cdot 2H_2O]$ is used and the temperature is melting point of glass <600°C. The typical bond strength achieved is ~9.8 kg cm^{-2}. (iv) *Thermocompression bonding*: it involves placing the two mirror-finished surfaces of the silicon in intimate contact at a high temperature. It leads to the development of *van der Waal forces* (intermolecular forces arising from polarization of molecules into dipoles or multipoles) between the two surfaces. In this case, the bond strength is weak.

3.4.6 Wafer Lapping

Back lapping is the thinning of semiconductor wafers by removing material from the backside, which is the unpolished or unprocessed face. The backside of the wafer is brought into contact with an abrasive slurry to remove material from this side. The slurry is a combination of lapping oil and silicon carbide (SiC) or aluminum oxide (Al_2O_3) particles.

3.4.7 Chemical Mechanical Polishing

Chemical mechanical polishing (CMP) is a process that is used for the planarization of semiconductor wafers. CMP takes advantages of the synergetic effect of both physical and chemical forces for polishing of wafers. Chemistry alone cannot planarize wafers because most chemical reactions are isotropic. On the other hand, mechanical grinding may achieve surface planarization but the surface damage is very high. Therefore, wafer polishing is done by

applying a load force to the back of a wafer while it rests on a pad. Both the pad and wafer are then counter rotated while a slurry containing abrasives and reactive chemicals is passed underneath.

3.4.8 Electroplating

Electroplated layers are of interest for several reasons. Patterned layers of nickel (Ni), nickel alloys, gold (Au) alloys, silver (Ag), and copper (Cu) are used to realize MEMS devices like micro switches, relays, valves, pumps, coils, and gyroscopes. The metal layers have to meet the highest requirements on homogeneity, as well as on mechanical, electric, or magnetic characteristics. Electroplating has to cope with specific effects influencing the conditions in the micrometer range.

Electroplating is defined as the deposition of a metal onto a metallic surface from a solution by electrolysis process. The electroplating process is well suited to make films of metals such as copper, gold, and nickel. The films can be made in any thickness from ~1 µm to >100 µm.

As an example, in copper plating, electroplating takes place by means of the reaction

$$Cu^{2+} + 2e \Leftrightarrow Cu \qquad (3.38)$$

Cu^{2+} in this equation represents an ion that is carried to the metal surface to be plated, known as the *cathode*, from the source of the metal being plated, known as the *anode*, which is made of copper. The ions are forced to the cathode by an external source such as a battery. The electrolytic solution is a salt of the metal being plated; in the case of copper plating, it is copper sulfate, $CuSO_4 \cdot 5H_2O$.

What are the primary requirements to be fulfilled by the substrate to be electroplated? It may be noted that electrodeposition requires the following: (i) electrical contact to the substrate when immersed in the liquid bath and (ii) further, the surface of the substrate must have an electrically conducting coating before the deposition can be done.

3.4.9 LIGA Process

What does LIGA stand for? LIGA is an acronym representing the main steps of the process steps involved, that is, deep XRL, electroforming (a metal forming process that forms thin parts through the electroplating process; it differs from electroplating in that the plating is much thicker and can exist as a self-supporting structure), and plastic molding (process used in manufacturing to shape materials), in German: Lithographie, Galvanoformung, Abformung (Figure 3.20). By deep XRL, structures of lateral design with high aspect ratios are produced, that is, with heights of up to 1 mm and lateral resolution down to 0.2 µm. The transparent carrier of the mask is a thin
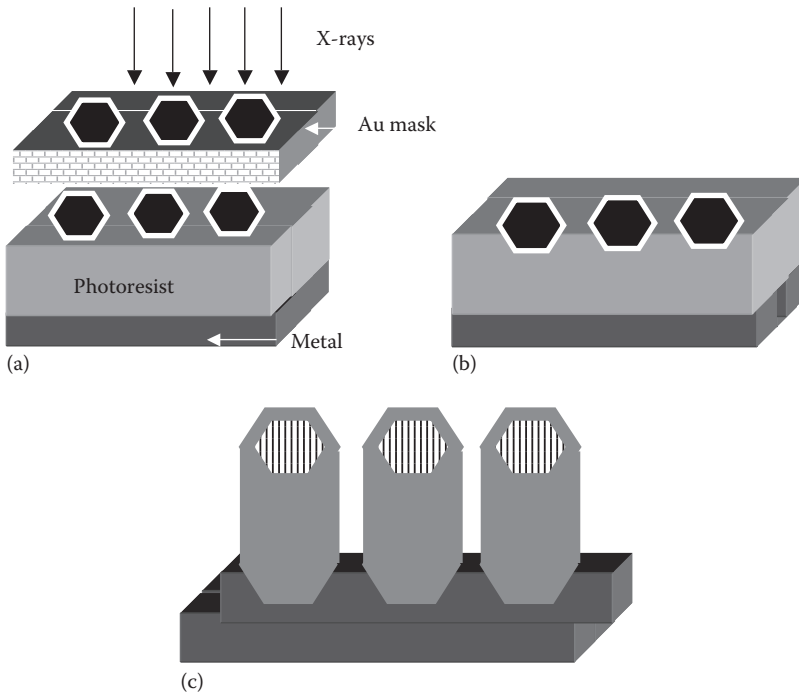
FIGURE 3.20
LIGA process: (a) exposure, (b) developing, and (c) electroplating and removing photoresist. (After Khanna, V. K., *Proceedings of IMS-2007, Trends in VLSI and Embedded Systems*, Punjab Engineering College, Chandigarh, 317, 2007.)

metal foil (e.g., titanium Ti, beryllium Be). The absorbers consist of a comparatively thick layer of Au. Synchrotron (cyclic particle accelerator) radiation is employed to transfer the lateral structural information into a plastic layer, generally polymethylmethacrylate (PMMA):$(C_5O_2H_8)_n$. Exposure to radiation modifies the plastic material in such a way that it becomes removable with a suitable solvent, leaving behind the structure of the unirradiated plastic as the primary structure.

The spaces generated by the removal of the irradiated plastic material are filled with metal using electroforming process. *What are the uses of micro/ nanostructures produced by LIGA process?* The metal micro/nanostructures produced by deep XRL and electroforming are used as molding tools for the production of faithful replicas of the primary structure in large numbers economically.

3.4.10 Micro-Injection Molding

It is a widely adopted procedure for fabrication of micro-parts or microstructured components from plastics, ceramics, and metals. The molding mass

is melted in the injection molding machine and then injected into a heated and frequently evacuated tool where it cools and solidifies into the final part. This tool is equipped with microstructured mold inserts, manufactured by mechanical micromachining (microcutting), laser micromachining, x-ray or UV lithography based on the LIGA process or by a combination of these and other processes. Depending on the component to be molded, the injection molding process is conducted either isothermally or with heating prior to injection and cooling prior to demolding. After cooling and opening of the tool, the injection-molded components are removed. This is done mostly by a handling device or robot. Nearly all thermoplastics (also known as thermo-softening plastics, are those becoming soft when heated and hard when cooled) as well as thermoplastic elastomers (polymers having the elastic properties of natural rubber: ELAST(IC)+[(POLY)MER) are suitable for micro-injection molding.

Noteworthy advantages of micro-injection molding are as follows: (i) cycle times are relatively short, (ii) components are fabricated in an integrated manner from several multicomponent materials, (iii) high degree of process automation is achieved, (iv) the process is applicable to a wide range of materials (plastics, ceramics, metals), and (v) commercially available equipment form the basis of system and tool technologies.

3.4.11 Hot Embossing and Electroforming

Hot embossing, is the stamping (making a distinctive mark or impression upon an object) of a pattern into a polymer that has been softened by raising its temperature above the *glass transition temperature* (the temperature at which the transition in the amorphous regions between the glassy and rubbery state occurs). The stamp used to define the pattern in the polymer is made by micromachining from silicon LIGA or other similar processes. *Hot embossing lithography* (*HEL*) has proved its potential for structuring resists with high aspect ratios by thermoplastic molding. It is an important fabrication process for topographic (having shapes and features on the surface) nanostructures to a lateral resolution as low as a few tens of nanometers. The process is parallel and has the implementational advantage of high throughput at a low cost. Figure 3.21 shows the process steps for fabrication of metal stamps with nanoscale fidelity (the quality of being faithful or loyal). By pressing a structured master onto a thermoplastic thin film, the latter is shaped down to sub-10 nm resolution in its viscous state.

Process parameters like temperature, pressure, vacuum, etc., are previously optimized to ensure complete filling of the cavities in the stamp with different aspect ratios and shapes. The polymer is hardened by cooling and then de-molded. For electroforming of hot embossed structures, a seed layer is deposited below the resist. Pattern transfer is done by removing the residual resist layer using oxygen RIE, which opens the seed layer windows for plating. Depending on the extent of electroplating, either the structure height is preserved or increased.
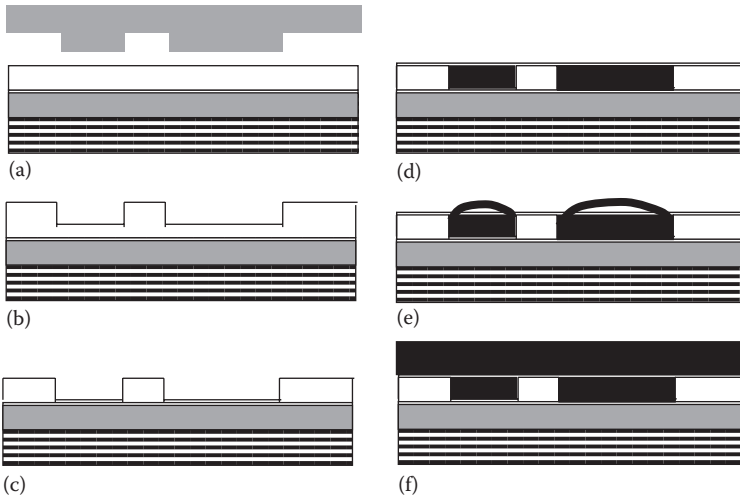
FIGURE 3.21
(a) Embossing, (b) de-molding, (c) dry etching, (d) electrofilling, (e) overplating, and (f) back-plating. (After Khanna, V. K., *Proceedings of IMS-2007, Trends in VLSI and Embedded Systems*, Punjab Engineering College, Chandigarh, 317, 2007.)

Daughter stamps are manufactured by extending the overplating to form a sizeable supporting base. Thus, a robust metal stamp, capable of withstanding repeated use and high mechanical loads, is constructed.

3.4.12 Combination of MEMS/NEMS and CMOS Processes

The *hybrid approach* in which the MEMS/NEMS devices and CMOS circuits are fabricated independently and connected together afterward is the most primitive method. There are three distinct approaches of fabrication of MEMS components along with interface electronics on the same chip: (i) *Preprocessing*: in which wells are etched up to a depth equal to the height required for fabrication of MEMS devices. Then MEMS devices are fabricated and protected with an encapsulating layer. The next step is CMOS circuit fabrication followed by the removal of the encapsulating layer to release the MEMS structures. (ii) *Postprocessing*: CMOS circuit is first fabricated and protected by a chemically resistant layer. Then micromachining steps are carried out avoiding high-temperature processing. (iii) *Combined processing* utilizing the CMOS layers for MEMS device fabrication.

3.5 Biochemistry Division

Biochemistry is the study of chemical and physicochemical processes of living organisms.

3.5.1 Surface Functionalization and Biofunctionalization of Nanomaterials

What is a functional group? A functional group is an atom or group of atoms within a molecule that is responsible for the characteristic properties of the molecule and reactions in which it takes part. *What does the functional group of a molecule tell us about?* It defines the structure of a family of compounds and determines the properties of the family, for example, methanol (CH_3OH), ethanol (C_2H_5OH), and isopropanol [$(CH_3)_2CHOH$] are all classified as alcohols since each contains a functional hydroxyl group. *What is meant by functionalization of a material?* Functionalization is the addition of functional groups onto the surface of a material by chemical synthesis methods.

What is functionalization of a material used for? (i) Functionalization is employed for modification of surfaces of industrial materials to achieve desired surface properties. Common examples are water-repellent coatings for automobile windshields and non-biofouling, hydrophilic coatings for contact lenses. (ii) In addition, functional groups are used to covalently link functional molecules to the surfaces of chemical and biochemical devices.

What does biofunctionalization mean? Biofunctionalization is the modification of a material (especially a nanomaterial) to infuse biological function in it. *Why is biofunctionalization done*? Biofunctionalization of gold, fluorescent, and magnetic nanoparticles, and carbon nanotubes is carried out as a routine procedure in biochemistry laboratories for fabricating various chemical and biosensors. These processes will be presented in relevant chapters in reference to the particular nanosensor being discussed. *What features should a biofunctionalization scheme take care of?* The appropriate biofunctionalization of nanoparticles clearly must address the features of colloidal stability, bioinertness, specificity with respect to target biomolecules, and low cytotoxicity.

Colloidal particles collide with each other due to the Brownian motion, convection, gravity, and other forces. Collisions may cause coagulation of the particles and destabilization of the colloid. Coagulation is the destabilization of colloids by neutralizing the electric charge of the dispersed phase particles, which leads to aggregation of the colloidal particles. Colloidal stability is achieved due to repulsion forces balancing the attraction forces. Two mechanisms are responsible for the colloidal stability: (i) Electrostatic stabilization of colloids is the mechanism in which the attractive van der Waals forces are counterbalanced by the repulsive Coulomb forces acting between the negatively charged colloidal particles. (ii) Polymeric stabilization of colloids involves the addition of polymeric molecules to the dispersion medium in order to prevent the aggregation of the colloidal particles. The polymeric molecules create a repulsive force counterpoising the attractive van der Waals force between particles approaching each other.

Does biofunctionalization affect (*a*) *the detection limit of bioanalytical systems,* for example, *biosensors or* (*b*) *its degree of selectivity and specificity?* It affects both

the parameters (a) and (b). Although the detection limit of the transducer surface is mainly determined by the detector element (transducer), the role of biofunctionalization cannot be neglected. For the same sensor, biofunctionalization incorporating nanoparticles gives lower limits of detection than done without these nano entities.

3.5.2 Immobilization of Biological Elements

What are the prominent biorecognition elements? Recognizable biorecognition elements include enzymes, for example, glucose oxidase (GO$_x$: an oxido-reductase that catalyzes the oxidation of glucose to hydrogen peroxide and ᴅ-glucono-δ-lactone: $C_6H_{10}O_6$), penicillase (an enzyme secreted by bacteria to inactivate the antibiotic penicillin), and urease (an enzyme that catalyzes the hydrolysis of urea into carbon dioxide and ammonia); antibodies, for example, anti-human Ig (Immunoglobulin) and anti-PSA (prostate-specific antigen); nucleic acids, for example, DNA and RNA. Other biological elements are cells, tissues, organelles, etc. *How is biofunctionalization accomplished?* Biofunctionalization is done by immobilizing so-called capture molecules (e.g., antibodies and enzymes) or ligands on the transducer surface, which are only able to bind to the corresponding analyte molecules ("lock-and-key" principle). As an alternative, sensors can be coated with biospecific polymers that are typically more cost efficient and chemically more stable than capture molecules; however, they may not be available for all analytes. *What main guidelines must be followed during biofunctionalization?* (i) When biofunctionalizing the transducer surface, the capture molecules or biospecific polymers must not be damaged, that is, the subsequent binding of analyte molecules must be guaranteed. (ii) Furthermore, the surface must be protected against binding of nonanalyte molecules contained in the sample medium (e.g., serum), for instance, by suitable intermediate layers, otherwise false-positive signals may be obtained vitiating the measurements. (iii) Besides, the immobilization procedure has to be chosen in accordance with the chemical environment provided by transducer and housing material as well as the detection principle, as this may limit the thickness of the sensitive layer. *What are the representative applications of preparing biosensor surfaces by biofunctionalization?* Mention may be made of detection of analyte concentrations, for example, concentration of low molecular compounds (such as penicillin in milk) and the concentration of disease markers in serum (e.g., breast cancer marker HER 2/ *neu* and inflammation marker c-reactive protein CRP).

What is the meaning of "enzyme immobilization"? Literally, "immobilization" is making incapable of being moved, that is, restricting the free movement of something. The enzymes are soluble in aqueous media. Immobilization means associating the biocatalysts with an insoluble matrix so that it is retained in a suitable reactor geometry for its economic reuse under stabilized conditions. Immobilization thus allows decoupling of the enzyme location from the flow of the liquid carrying the reagents and products.

What are the main classes of methods for immobilization of enzymes? A large number of techniques are available for the immobilization of enzymes or cells on a variety of natural and synthetic supports. The choice of the support as well as the technique depends on the nature of the enzyme and the substrate and its ultimate application. Therefore, it is not possible to recommend any single universal means of immobilization. It can only be commented that the search must continue for matrices that provide facile, secure immobilization of enzyme along with good interaction with substrates, and which conform in shape, size, and density to the intended use. A combination of one or more of these techniques is also sometimes applied. Methods used for the immobilization of enzymes fall into four main categories:

1. *Physical adsorption of enzyme on an inert carrier*: This is perhaps the simplest technique of preparing immobilized enzymes, one that does not grossly or coarsely alter the activity of the bound enzyme. It relies on nonspecific physical interaction between the enzyme protein and the surface of the matrix. It is achieved by mixing a concentrated solution of enzyme with the solid. Usually no reagents are necessary. Only a minimum of activation steps is required so that the adsorption is cheap, easily carried out, and is less disruptive to the enzyme protein. Binding is secured mainly by hydrogen bonds, multiple salt linkages, and van der Waals forces. *But these are frail forces? What is the effect of using these weak forces for immobilization on the enzyme function?* As these binding forces are inadequate, it is not astonishing that very often the protein desorbs due to changes in temperature, pH, ionic strength, etc. Also, due to the deficiency in binding strength, the properties of the enzyme are sometimes altered by further adsorption of other proteins or substances during use of the immobilized enzyme. This variation of properties severely impairs the device function.

2. *Inclusion of enzyme in the lattice of a polymerized gel*: The hydrophilic (readily mixing with or wetted by water) matrix is polymerized in an aqueous solution of the enzyme and the polymeric mass is broken up to the desired particle size. As no bond is formed between the enzyme and the polymer matrix, disruption of the protein molecules does not take place. However, free radicals generated during polymerization affect the activity of entrapped enzymes. Further, the method is unsuitable for those enzymes that act on macromolecular substrates, such as ribonuclease (commonly abbreviated Rnase, is a type of nuclease that catalyzes the degradation of RNA into smaller components), trypsin (enzyme that acts to degrade protein; it is often referred to as a proteolytic enzyme, or proteinase), and dextranase (an enzyme that catalyzes the endohydrolysis of 1,6-[[alpha]]-glucosidic linkages in dextran). This is because only low-molecular-weight substrates are able to diffuse rapidly to the enzyme.

3. *Cross-linking of the protein with a bifunctional reagent*: *What is a bifunctional reagent?* It is a reagent with two reactive groups, usually at opposite ends of the molecule, that are capable of reacting with and thereby forming bridges between side chains of amino acids in proteins. This helps in identifying the locations of naturally reactive areas within proteins.

   In this enzyme immobilization technique, intermolecular cross-linking of the protein is carried out. This linking is done either to other protein molecules or to functional groups on an insoluble support matrix. Enzymes are immobilized through chemical cross-linking using homo- as well as heterobifunctional cross-linking agents. Two bifunctional reagents are glutaraldehyde, as an amine-reactive homobifunctional cross-linker (CL) and dimethylsuberimidate (DMS), a bifunctional imidoester. *What is meant by homo- and heterobifunctional cross-linking agents?* CLs are either homo- or heterobifunctional reagents with identical or nonidentical reactive groups, respectively, permitting the establishment of inter- as well as intramolecular cross-linkages. *Among the cross-linking agents, which substance has been widely used and why?* Glutaraldehyde [$CH_2(CH_2CHO)_2$], which interacts with the amino groups through a base reaction, has been extensively used in view of its low cost, high efficiency, and stability.

   The enzymes or the cells have been normally cross-linked in the presence of an inert protein like gelatine (colorless or slightly yellow, tasteless, transparent, brittle water-soluble protein derived from collagen, and formed by boiling the specially prepared skin, bones, and connective tissues of animals), albumin (water-soluble protein found in egg white, milk, blood, etc.), and collagen (a fibrous protein; the main structural protein in animal extracellular matrix and connective tissue, forming molecular cables that strengthen the tendons and vast, resilient sheets that support the skin and internal organs). As some of the protein material is inevitably always acting as a support, there is relatively low enzyme activity. Cross-linking is therefore best used in conjunction with one of the other methods.

4. *Covalent binding of enzyme to a reactive insoluble support*: This acts through formation of covalent bonds between the enzyme and the support matrix. The binding reaction must be performed under conditions that do not cause loss of enzyme activity. Also, the active site of the enzyme must be unaffected by the reagents used. *Which functional groups of proteins are suitable for covalent binding?* The functional groups of proteins suitable for covalent binding under gentle conditions are the alpha amino groups (attached to the carbon atom immediately adjacent to the carboxylate group [the number two, or α–carbon]; generic formula $H_2NCHRCOOH$, where R is an organic substituent) of the chain and the epsilon amino groups (the amino

acid lysine has a positively charged amino group in its side chain located at the epsilon-carbon and hence called the epsilon-amino group) of lysine (α-amino acid well known for its antiviral properties, present in protein and essential in the diet of vertebrates, with the chemical formula $HO_2CCH(NH_2)(CH_2)_4NH_2$) and arginine (1 of the 20 amino acids that serve as the building blocks in protein, benefits hard training athletes; not essential to the diet, but can be made by the body from other substances); the alpha carboxyl group of the chain end and the beta and gamma carboxyl groups of aspartic (amino acid present in many proteins, important in animal metabolism and acting as a neurotransmitter) and glutamic (a naturally occurring amino acid found in several proteins) acids; the phenol ring of tyrosine (a hydrophilic amino acid present in many proteins and participating in synthesis of hormones); the thiol group (a sulfhydryl group –SH) of cysteine (sulfur-containing amino acid: $C_3H_7NO_2S$, which is a constituent part of many enzymes and essential in human diet); the hydroxyl groups of serine (a hydrophilic amino acid in proteins) and threonine (an amino acid essential for growth: $HO_2CCH(NH_2)$ $CH(OH)CH_3$); the imidazile group ($C_3H_4N_2$) of histidine (an essential amino acid playing a role in maintaining health, vitality, and well-being: $C_6H_9N_3O_2$) and the indolyl group (C_8H_6N derived from indole C_8H_7N by removal of a hydrogen atom from a ring atom) of tryptophan (an amino acid formed from proteins during digestion by the action of proteolytic enzymes, needed for normal growth in infants and for nitrogen balance in adults: $C_{11}H_{12}N_2O_2$).

Does covalent immobilization affect the functional groups of enzymes responsible for catalytic action? Enzymes are covalently linked to the support through the functional groups in the enzymes, which are not essential for the catalytic activity. It is often advisable to carry out the immobilization in the presence of its substrate or a competitive inhibitor (molecules that bind to enzyme, alter the catalytic action of the enzyme and consequently slow down, or in some cases, stop catalysis; the inhibitor has a similar shape to the usual substrate for the enzyme, and competes with it for the active site but once it is attached to the active site, nothing happens to it because it does not react—essentially, it just gets in the way) so as to protect the active site. The covalent binding should also be optimized so as not to alter its conformational flexibility; "conformation" refers to the structural arrangement or the way a substance is formed. Some of these problems, however, can be obviated by covalent bonding through the carbohydrate moiety when a glycoprotein (any of a class of compounds consisting of a protein in combination with a carbohydrate) is concerned.

What are the favorable features for this immobilization strategy? (i) Due to the wide variety of binding reactions, and insoluble carriers with functional groups

capable of covalent coupling, or being activated to give such groups, this is a generally applicable method of insolubilization (to make insoluble). (ii) Like cross-linking, covalent bonding provides stable, insolubilized enzyme derivatives that do not leach or percolate enzyme into the surrounding solution.

What are the criteria for selection of a particular method of enzyme immobilization amongst those described in the preceding paragraphs? (i) Knowledge of active sites of particular enzymes enables methods to be chosen that avoid reaction with the essential groups therein. One must carefully choose a method of attachment aimed at reactive groups outside the active catalytic and binding site of that enzyme. As an alternative, these active sites can be protected during attachment as long as the protective groups are removable without loss of enzyme activity. In some cases, this protective function is fulfilled by a substrate of the enzyme or a competitive inhibitor. (ii) Several vital roles are played by the surface on which the enzyme is immobilized. These include retention of tertiary (referred to as third stage, third level is third in order or rank) structure in the enzyme by hydrogen bonding or the formation of electron transition complexes. Retention of tertiary structure may also be a crucial factor in maximizing thermal stability in the immobilized state.

How does the microenvironment imposed upon enzymes by their supporting matrices and by the products of their own action affect their properties?

1. The first important property is stability. Depending on whether the carrier provides a microenvironment capable of denaturing (changing the natural qualities of a protein by heating, acidity, etc.; a structural change caused by extreme conditions) the enzyme protein or of stabilizing it, the stability of the enzyme either increases or decreases. To reduce inactivation by auto digestion of proteolytic enzyme (enzyme [protease] that helps to digest the proteins in food to liberate the amino acids needed by the body), the enzyme molecules are isolated from mutual attack by immobilizing them on a matrix.

2. The second property of enzyme is its specific activity. Generally, the specific activity of an enzyme decreases upon insolubilization. This is due to denaturation of the enzyme protein caused by the coupling process. It may arise from physical and chemical character of the support matrix alone, or may result from interactions of the matrix with substrates or products involved in the enzymatic reaction. It must not be forgotten that the immobilized enzyme is residing in a drastically different microenvironment from that existing in free solution.

3.5.3 Protocols for Attachment of Antibodies on Sensors

Sensors such as microcantilevers require biofunctionalization for their operation. Several protocols have been developed for this purpose. As an

example, Kale et al. (2007) have described a detailed process to biofunction-alize silicon nitride (Si_3N_4) cantilevers. This process consists of the following steps: (i) The cantilever was subjected to sulfochromic acid (H_2SO_4/CrO_3) treatment for 10 min producing silanol (Si–OH group) sites, and dried by heating in vacuum. (ii) The cantilever was immersed in 1% silane for 5 min. Silanes (compounds of silicon and hydrogen of the formula Si_nH_{2n+2}) and other monomeric silicon compounds are able to bond inorganic materials such as glass, metals, and metallic oxides to organic resins. The adhesion mechanism is due to two groups in the silane structure: (a) The $Si(OR_3)$ por-tion reacts with the inorganic reinforcement, while (b) the organofunctional (vinyl: $-CH=CH_2$; amino: ($-NH_2$); epoxy: an oxygen atom joined by single bonds to two adjacent carbon atoms, thus forming the three-membered epoxide ring, etc.) group reacts with the resin. The coupling agent is applied to the inorganic materials as a pretreatment and/or added to the resin.

After silane application to the cantilever, the excess silane was removed by rinsing in ethanol, and the cantilever was heated at 110°C for 10 min in argon atmosphere. (iii) The cantilever was dipped in 1% aqueous solution of glutaraldehyde [$CH_2(CH_2CHO)_2$] for 30 min. The linker molecules got bound to surface amine sites (molecules that contain the functional group $-NH_2$). Thus, free aldehydes were left on the surface to which biomolecules can attach. (iv) On the modified cantilever surface, human immunoglobu-lin (HigG) antigens were immobilized by incubating them on the canti-lever surface for 1 h. Loosely adsorbed biomolecules on the surfaces were subsequently removed by washing with a detergent solution and rinsing with PBS (phosphate buffered saline: $KCl + KH_2PO_4 + NaCl + Na_2HPO_4$). (v) FITC (Fluorescein-5-Isothiocyanate: $C_{21}H_{11}NO_5S$)-tagged goat anti-HigG (1 mL mL^{-1} in PBS) was allowed to react with the sample surfaces for 1 h, and the cantilever die was rinsed; FITC is a commonly used fluorescent dye. It is an amine-reactive fluorescein derivative suitable for protein labeling. (vi) The cantilever die was inspected under a fluorescence microscope. If the FITC-tagged antibodies were attached to the corresponding antigens, they became visible under the fluorescence microscope. This microscope is a variation of the compound laboratory light microscope, which is arranged to admit ultraviolet, violet, and sometimes blue radiations to a specimen, which then fluoresces. Basically, the microscope irradiates the specimen with a desired and specific band of wavelengths and then separates the much weaker emit-ted fluorescence from the excitation light. Thus, the sample to be studied is itself the light source. First, the microscope has a filter that only allows radiation with the desired wavelength that matches the fluorescing mate-rial. To become visible, the emitted light is separated from the much brighter excitation light in a second filter. Here, the fact that the emitted light is of lower energy and has a longer wavelength is used. The fluorescing areas are observed in the microscope and shine out against a dark background with high contrast. In a properly configured microscope, only the emission light reaches the eye or detector so that the resulting fluorescent structures are

superimposed with high contrast against a very dark (or black) background. The limits of detection are generally governed by the darkness of the background, and the excitation light is typically several 10^5–10^6 times brighter than the emitted fluorescence.

It is necessary to mention here that FITC is a small organic molecule that is typically conjugated to proteins via primary amines, that is, lysines. Generally, 3–6 FITC molecules are conjugated to each antibody; higher conjugations cause solubility problems as well as internal quenching and reduced brightness. Thus, an antibody is usually conjugated in several parallel reactions to different amounts of FITC, and the resulting reagents are compared for brightness and background adherence to decide the optimal conjugation ratio. Fluorescein is typically excited by the 488 nm line of an Ar laser. The emission is collected at 530 nm.

3.5.4 Functionalization of CNTs for Biological Applications

How are CNTs functionalized for biological applications? For biological uses, CNTs are functionalized by attaching biological molecules such as lipids, proteins, biotins, etc., to them. In this way, it is possible to solubilize and disperse carbon nanotubes in water, thus opening the path for their facile manipulation and processing in physiological environments. Then they can usefully mimic certain biological functions, such as gene therapy and drug delivery. In biochemical and chemical applications, carboxylic acid poly-amino benzoic sulfonic acid, polyvinyl alcohol, amino acid derivatives, halogens, and compounds have been used to functionalize CNTs.

3.5.5 Water Solubility of Quantum Dots

For using in biological applications, quantum dots need to be water soluble. TOPO [tri-*n*-octylphosphine oxide: $OP(C_8H_{17})_3$]-stabilized quantum dots, show hydrophobic (aversion to water) surface properties (Murcia and Naumann 2005). *What surface modification schemes have been applied for dispersing TOPO-stabilized QDs in water?* To disperse TOPO-stabilized quantum dots in aqueous solution, several surface modification strategies have been pursued: (i) A common approach is to synthesize quantum dots in TOPO and replace the hydrophobic TOPO layer with bifunctional molecules containing thiol and hydrophilic moieties separated by a molecular spacer. The thiol groups bind to the CdSe or ZnS surface, while the hydrophilic moieties diverge from the surface of the corresponding semiconductor. Unfortunately, thiols bind less strongly to ZnS than to Au, which leads to a dynamic equilibrium between bound and unbound thiols. This behavior degrades the long-term water solubility of ZnS-capped quantum dots. To shift the equilibrium toward bound moieties, monothiols have been replaced with molecules containing more than one thiol group. (ii) Another stabilization concept is to enhance binding via surface cross-linking of bound molecules. On the basis

of this concept, ZnS-shelled quantum dots have been made water soluble by adding a silica shell to the nanoparticles by using alkoxysilanes ($R_nSiX_{(4-n)}$) where R is a nonhydrolyzable organic moiety that can be either an alkyl, aromatic, organofunctional, or a combination of any of these groups) during the polycondensation (a chemical condensation of a monomer having two functional groups leading to the formation of a polymer, that is, bifunctional or multifunctional monomers react to form first a dimmer, then a trimer, longer oligomer, and eventually long-chain polymer, a polycondensate, along with the releasing of water or a similar simple substance). Two types of silanes have been used to stabilize quantum dots in aqueous solution. The first category includes silanes whose surface functional groups are positively or negatively charged at neutral pH. The second type includes silanes with poly(ethylene glycol) [$C_{2n+2}H_{4n+6}O_{n+2}$] chains. (iii) TOPO-coatings also can be made water soluble without their replacement by adding amphiphilic molecules (chemical compounds possessing both hydrophilic [*water-loving*] and lipophilic [*fat-loving*] properties) such as lipopolymers (lipids with a polymer chain covalently attached to the lipid) or amphiphilic diblock copolymers (block copolymers with two distinct blocks; copolymers are those derived from two or more monomeric species and block copolymers comprise two or more homopolymer subunits linked by covalent bonds), whose hydrophobic moiety stabilizes the TOPO-coating via hydrophobic forces and whose hydrophilic moiety is exposed to the solvent environment, guaranteeing water solubility. The last approach gives the advantage of not exposing the sensitive surface of the quantum dot during a surface exchange step.

3.5.6 Low Cytotoxicity Coatings

Successful application of fluorescent nanoparticles in a biological environment requires not only high dispersion stability and the suppression of nonspecific biomolecule adsorption, albeit they should infallibly show low cytotoxicity (producing a toxic effect on cells) (Murcia and Naumann 2005). Particularly, this issue is of great consequence for applications involving quantum dots where toxic heavy metal ions may be released. If properly passivated, quantum dots give the impression of insignificant cytotoxicity. *What are the approaches followed for toxicity reduction of QDs?* (i) One approach is simply to passivate the quantum dot surface with binding ligands (atoms or molecules attached to a central atom, usually a metallic element, in a coordination or complex compound). Interestingly and importantly, such a passivation has a dual role because it not only lowers the cytotoxicity of QDs, but provides the bonus of enhancing their quantum yields. (ii) Quantum dots can be further passivated by adding a protecting semiconductor shell. Such coatings also considerably lower the cytotoxicity of CdSe quantum dots, but do not completely eradicate the problem. (iii) Surface silanization is another promising approach to suppressing surface oxidation. The stability of the coating is provided by cross-links within the

siloxane (compound that has alternate silicon and oxygen atoms composed of units of the form R_2SiO, where R is a hydrogen atom or a hydrocarbon group) shell. Decisively, such shells are quite stable in a biological environment. Furthermore, CdSe/ZnS quantum dots with a protective shell of cross-linked silica significantly reduce the release of Cd^{2+}, thus diminishing the cytotoxicity of fluorescent nanoparticles. (iv) Fascinatingly, the addition of a bovine serum albumin (BSA) layer further lessens the cytotoxicity of CdSe/ZnS quantum dots. The BSA, added to the mercaptoacetic acid-functionalized quantum dot surface using EDC coupling, acts as a diffusion barrier for O_2 molecules.

3.6 Chemistry Division

Surface modification is applied in the fabrication of many chemical sensors for gas detection. A few examples will be cited in the following text.

3.6.1 Nanoparticle Thin Film Deposition

Thin films composed of nanoparticles have emerged as useful chemical sensor platforms. Numerous other techniques exist for the deposition of nanoparticle films: (i) controlled pulling, (ii) the use of charged polymers and charged nanoparticles, or (iii) slow incubation of thiol-capped nanoparticles in dithiol (compound having two thiol groups) solution. In a report (Childs et al. 2005), a technique is described to deposit nanoparticle films of different thickness on prepared electrodes in a reproducible manner. The nanoparticles used for these studies employed 5 nm gold and 2 nm platinum particles. The process steps are as follows: (i) Substrates were first silanized with a solution of tetrakis(dimethylamino)silane, $Si[N(CH_3)_2]_4$, in toluene (C_7H_8). (ii) The remaining amino functionalities were displaced with 1,8-octanedithiol (ODT) [$HS(CH_2)_8SH$] from a toluene solution to give a thiolated (reacted with a thiol) surface capable of binding nanoparticles. (iii) The substrate was then incubated in a solution of dodecylamine ($C_{12}H_{27}N$)-capped metal nanoparticles dissolved in toluene or hexane. (iv) The film thickness was increased via alternating exposure to solutions of bifunctional cross-linking molecules (containing two reactive groups providing a means of covalently linking two target groups) and nanoparticles.

3.6.2 Polymer Coatings in Nano Gas Sensors

Here a nanomaterial, for example, a CNT, is coated with specific polymers to make the CNT sensitive to a particular gas. For certain gases and vapors, pristine SWCNTs do not respond at all, and in those cases, coating or doping

of the nanotubes is able to elicit a signal. Li et al. (2006) coated SWCNTs deposited on an interdigitated electrode pattern with polymers such as chlorosulfonated polyethylene (CSPE: synthetic rubber) and hydroxypropyl cellulose (HPC: a partially substituted, water-soluble cellulose ether) to make them sensitive to chlorine (Cl_2) and hydrochloric acid (HCl) vapor, respectively. *What was the coating procedure followed?* 0.804 mg of chlorosulfonated polyethylene was dissolved in a solvent of 25 mL of tetrahydrofuran (THF: C_4H_8O) as a coating solution for sensing Cl_2. Hydroxypropyl cellulose (0.791 g) was dissolved in a solvent of 25 mL of chloroform ($CHCl_3$) for coating the nanotubes to detect HCl. In each case, a 5 nL aliquot of polymer solution was drop deposited onto the SWCNT network to get the SWCNTs coated.

3.6.3 Metallic Nanoparticle Functionalization of Si Nanowires for Gas Sensing Applications

Chen et al. (2007) fabricated Si nanowires (NWs). Pd nanoparticles were deposited on Si NWs for H_2 sensing. *How was this done?* First, the oxide covering on the as-grown Si NWs was etched away by dipping in a 5% HF aqueous solution for 10 min. After rinsing with distilled water, Pd particles were deposited on the surface of Si NWs by immersion in a saturated palladium chloride ($PdCl_2$) solution for 10 min. Consequently, the Pd^{2+} ions were reduced and Pd metal particles were deposited onto the surface of Si NWs in accordance with the following chemical equation:

$$Pd^{2+} + 2Si - H \rightarrow Pd + H_2 + 2Si \qquad (3.39)$$
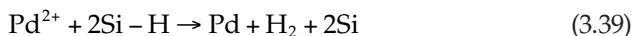
The Pd-modified Si NWs thus obtained were dispersed on a silicon wafer with a 300 nm oxide layer. For contacts, gold electrodes were deposited onto the sample by thermal evaporation and current–voltage characteristics were measured with a two-probe system.

Upon exposure to 5% hydrogen, the current signal of the sensor increased by about 20 times. The response time was 3 s only, which was appreciably faster than that of the macroscopic Pd wire sensor. This device will be further elaborated in the chapter on chemical nanosensors.

3.7 Nanosensor Characterization Division

This division owes the responsibility to test and evaluate the developed nanosensors in an unbiased manner. The role of this division cannot be underestimated because proper evaluation of new nanosensors, a critical assessment of their properties and provision of feedback to the device design and processing personnel, is essential for improvements in their nanosensor

characteristics. As a first step, the nanosensor in question should be classified as belonging to which category: physical, chemical, or biological. Accordingly, it is diverted to the appropriate characterization facility. As the testing methods for these nanosensors are different, it is abundantly clear that the characterization facility will have rooms in physics, chemistry, and biology divisions to undertake the testing jobs. Obviously, a blood glucose sensor cannot be tested in a physics lab. However, its physical transducer can be tested there. Apart from general characterization, the above facility also looks into the unique characterization tests for individual nanosensors and upgrades the instrumentation periodically as deemed necessary by the requirements.

3.8 Nanosensor Powering, Signal Processing, and Communication Division

Most of the issues in this field are presently research challenges that will be addressed by futuristic nanosensor devices and systems (Akyildiz and Jornet 2010). Therefore, the scenario is more envisioned and imaginary than realistic.

3.8.1 Power Unit

3.8.1.1 Lithium Nanobatteries

Nanomaterials can be used to manufacture nanobatteries with high power density. *What are nanobatteries?* These are batteries fabricated using nanoscale technology. In manufacturing batteries, nanotechnology offers the following benefits: (i) It increases the available power from a battery and decreases the time required to recharge a battery. *How is this possible?* This is achieved by coating the surface of an electrode with nanoparticles, hence increasing the surface area of the electrode, and thereby allowing more current to flow between the electrode and the chemicals inside the battery. (ii) It increases the shelf life of a battery by using nanomaterials to separate liquids in the battery from the solid electrodes under the condition that no current is being drawn from the battery. Such separation prevents the low-level discharge occurring in a conventional battery, which increases the shelf life of the battery severely.

Lithium ion (or Li-Ion) batteries are one of the most common types of batteries used in consumer electronic products ranging from mobile phones to laptop computers. The origins of lithium batteries date back to the beginning of the twentieth century. Lithium nanobatteries were constructed using alumina (Al_2O_3) membranes having pores 200 nm in diameter (Vullum and

Teeters 2005, Vullum et al. 2006). Each one of these pores was filled with Poly Ethylene Oxide (PEO: $(OCH_2CH_2)_n$ where n represents the average number of oxyethylene groups) and lithium triflate (CF_3LiO_3S), that is, (PEO + lithium triflate) electrolyte and capped with cathode material. V_2O_5 ambigel confined in the pores served as cathodes for individual nanobatteries. The anode was made of lithium. Thus, each pore became an effective nanobattery. Individual nanobatteries in the arrays were characterized by charge/discharge tests using the cantilever tip of an AFM to make electrical contact with the 200 nm cathodes of the nanobatteries. The volumetric capacity for each individual nanobattery was ~23–30 μA h^{-1} cm^{-2} μm^{-1}, which is confirmatory of its promise for delivering power to nanodevices. These nanobatteries were charged and discharged and had good capacities, proving that these small batteries could function as minute power systems.

Lithium batteries could function as viable, miniaturized power sources for future nanodevices.

One of the most important aspects of this nanobattery technology compared to existing battery types is its safety. Since each nanobattery comprises thousands of small batteries, even if one of these small batteries has a short circuit and fails, the entire battery can keep functioning, losing only a very small amount of power. Similar damage to a conventional Li-ion battery could result in substantial loss of power or a complete malfunction and in extreme cases even fire or explosion.

3.8.1.2 Self-Powered Nanogenerators

The operation of these devices is based on the transformation of the following types of energy into electrical energy (Wang 2008): (i) *Mechanical energy*: sources of this energy are the human body movements, or muscle stretching like walking, vibrations, jerks, etc. (ii) *Vibrational energy*: this is generated by acoustic waves or structural vibrations of buildings, such as the sound of human speech, vehicles, or other noise. (iii) *Hydraulic energy*: it is produced by body fluids or the blood flow.

The conversions of these energies are effected by piezoelectric effect in zinc oxide (ZnO) NWs. When these NWs are subjected to mechanical deformations, such as upon bending, a potential difference appears in the NWs.

3.8.1.3 Energy Harvesting from the Environment

Nanotubes and nanocantilevers can be designed to absorb vibrational energy at specific frequencies. A nonlinear oscillator can harvest energy from a wide spectrum of vibrations or even mechanical and thermal noise. Besides mechanical, vibrational, or hydraulic energy (derived from the force or energy of moving water), it is also possible to gather energy from electromagnetic waves in the nanoscale, for example, an NEMS-based resonator can be applied to convert electromagnetic radiation into vibrational energy. This

can be transformed into electricity using ZnO NWs. Alternatively, nanoscale "rectennas" can be developed using CNTs. These are rectifying antennas that convert electromagnetic waves into DC electricity that can be used to supply power to nanosensors.

A simple rectenna element consists of a dipole antenna with a Schottky diode placed across the dipole elements. This diode rectifies the AC current induced in the antenna by the microwaves, to produce DC electricity. The reason for choosing Schottky diodes is that they have the lowest voltage drop and highest speed and therefore waste the least amount of power due to conduction and switching. Large rectennas comprise an array of many such dipole elements.

3.8.1.4 Synthetic Chemical Batteries Based on Adenosine Triphosphate

Adenosine triphosphate (ATP) is the energy source of cells and living organisms, which can be obtained by chemical reactions in the nanoscale emulating cell respiration. It is considered by biologists to be the energy currency of life. ATP batteries imitating the behavior of mitochondria are an alternative energy source for bionanodevices.

3.8.2 Signal Processing Unit

Enabling of nanoscale processors is facilitated by the development of different forms of smaller FET transistors. Graphene (an allotrope of carbon)-based transistors are not only smaller but are expectedly faster. As graphene shows almost ballistic transport of electrons, the electrons traverse larger distances without their back-scattering permitting the fabrication of swifter switching devices. Additionally, the reduction of the channel length contributes to a faster response of the transistor. Theoretically, the switching frequencies of graphene-based transistors are predicted around few hundreds of terahertz, which are higher than any existing silicon FET. However, irrespective of the specific approach adopted to design these nanotransistors, the main challenge will be their integration in future processor architectures.

3.8.3 Integrated Nanosensor Systems

The capability of a nanosensor is noticeably augmented by incorporating the signal processing circuit and the nanosensor on the same chip. Such a system is called an integrated nanosensor system (INS). Once the performance of a nanosensor has been optimized individually, this kind of integration will be widely practiced.

Integrated nanosensors can go far beyond simple nanosensors providing features such as standard interfaces, self-testing, fault tolerance, and digital compensation to extend overall system accuracy, dynamic range, and reliability. Another aspect that has been neglected in considering

integrated nanosensors is that a variety of new nanosensors, hitherto not practical, may be feasible. There exists a wide unexplored range of possibilities for combining digital signal processing and instrumentation techniques.

3.8.4 Wireless Nanosensor Networks

Why are wireless nanosensor networks (WNSNs) *a very exciting research area?* This is because they are the enabling technology of many applications that will impact our society and change our daily lives, encompassing healthcare to homeland security and environmental protection. The technology of wireless sensor networks helps to run factories, optimize widely spread processes, monitor the weather, detect the spreading of toxic gases in chemical industries, and even provide precious extra time in advance of tornados and earthquakes. The widespread use of wireless sensor networks driven by the concept of distributed sensing and computing of indoor and outdoor environment has revolutionized the present state of environmental protection and control.

Two main methods for communication are envisioned in the nanoscale: (i) *molecular communication* defined as the transmission and reception of information encoded in molecules and (ii) *nanoelectromagnetic communication* dealing with the transmission and reception of electromagnetic radiation from components utilizing novel nanomaterials. Electromagnetic communication among nanosensors will be aided by the development of nanoantennas and the corresponding electromagnetic transceiver. To seek the nanoantenna goal (Hagerty et al. 2004), the following approaches are likely: (i) More accurate models for nanoantennas based on nanotubes and nanoribbons need to be formulated by providing their specific bands of operation, radiation bandwidth, and radiation efficiency. All these parameters will determine the communication capabilities of nanosensor devices. (ii) Novel nanoantenna designs and radiating nanostructures must be put forward by exploiting the properties of nanomaterials and new manufacturing techniques. (iii) A new antenna theory must be framed by considering the quantum effects observed in the nanoscale.

The EM transceiver of a nanosensor device will embed the essential circuitry to perform baseband processing, frequency conversion, filtering, and power amplification of the signals to be broadcast or those received from the free-space through the nanoantenna. Toward this aim, the suggestions are as follows: (i) More accurate models for the electromechanical nanotransceiver are required by accounting for the radiation bandwidth and energy efficiency of the complete process. (ii) The noise in reception needs to be characterized by recognizing which types of noise affect the electromechanical unit and how they influence the demodulated signal. (iii) New nanoreceiver architectures able to support more advanced, robust or bandwidth efficient modulations should be developed.

3.9 Discussion and Conclusions

A nanosensor fabrication facility was conceptualized. The pluridisciplinary character of the laboratory was envisaged. Not only the fabrication facility, the characterization equipment also need inputs bordering and cutting across various fields of science and technology. A strong infrastructural support is required to keep these facilities in running condition. A nanosensor designer chooses the nanomaterial and accordingly selects the enabling fabrication technology to realize the device from amongst the wide-ranging options described.

Review Exercises

3.1 Nanoparticles of noble metals are characterized by the presence of bright colors. What is the underlying cause of these colors?

3.2 Is the position of surface plasmon band dependent on the nanoparticle size? What is the role of nanoparticle shape in deciding the position of this band? What is the effect of interparticle distance on the band position?

3.3 Are the following statements true or false?

    (1) Larger particles show absorption that is more red-shifted as compared to the smaller ones.

    (2) Aggregates of nanoparticles have plasmon oscillations that become red-shifted as compared to the individual nanoparticles, a phenomenon widely used in chemical and biological sensors.

3.4 (i) Do Si NPs exhibit size-dependent tunable light emission? (ii) Do Si nanoparticles exhibit a low inherent toxicity in comparison with the heavy elements of several other types of semiconductor quantum dots?

3.5 Unlike bulk silicon, the emission from nanoscaled silicon can be attributed to radiative recombination of carriers confined in Si NPs. Do you agree?

3.6 Define the large-scale production of CNT in terms of number of tons/annum per plant. Mention three reasons justifying why CVD has become the most important commercial approach for manufacturing CNTs. Name two other commonly used methods of CNT synthesis.

3.7 What are the control parameters for CVD nanotube synthesis? By what technique are large arrays of well-aligned CNTs synthesized?

3.8 Why should agglomerated metal oxide nanoparticles be minimized for gas sensor applications? A tin oxide nanoparticle gas sensor takes a time of 110 s to attain 63.2% of the initial difference in electrical resistance before and after changing the gas environment. What is the time constant of the sensor?

3.9 Do silver nanoparticles exhibit antimicrobial activity? Can these nanoparticles be used in medicine to reduce infections as well as to prevent bacteria colonization on prostheses? Describe one method of synthesizing silver nanoparticles.

3.10 A primary process for semiconductor nanofabrication is RIE, which uses chemical and physical processes to etch away desired materials. Compare RIE with wet etching in terms of anisotropy, uniformity and control, and etch selectivity. Differentiate between RIE, sputter etching, and vapor phase etching.

3.11 In the broadest sense, CVD involves the formation of a thin solid film on a substrate material by a chemical reaction of vapor-phase precursors. How do you distinguish it from PVD processes, such as evaporation and reactive sputtering?

3.12 Compare LPCVD and PECVD processes in terms of film deposition conditions and resultant film quality.

3.13 Name two processes of junction formation, that is, transition from P- to N-type or vice versa. What is meant by substitutional and interstitial diffusion of impurities in a semiconductor?

3.14 How are phosphorous pre-deposition and drive-in cycles carried out in a semiconductor fabrication facility? How are boron pre-deposition and drive-in performed? Write the pertaining chemical equations.

3.15 Boron and phosphorus are the basic dopants of most ICs. Arsenic and antimony, which are highly soluble in silicon and diffuse slowly, are used before epitaxial processing or as a second diffusion. What are the difficulties faced in using gallium, indium, and aluminum as P-type dopants in IC industry?

3.16 Phosphorus is useful not only as an emitter and base dopant, but also for gettering fast-diffusing metallic contaminants, such as Cu and Au, which cause junction leakage current problems. Thus, phosphorus is indispensable in VLSI technology. Elaborate these remarks.

3.17 Gold diffuses very fast in silicon. What is the application of gold in silicon technology?

3.18 What are positive and negative resists used in photolithography? Differentiate between soft and hard-baking steps.

3.19 Most wet etchants are isotropic, meaning that they etch silicon equally in all directions. This causes a well-known phenomenon

in the microelectronics industry called *undercut*. How is undercutting avoided in deep RIE?

3.20 Distinguish between pin-through-hole and surface mount technologies in IC packaging. Compare plastic and ceramic packaging materials.

3.21 What are the two main types of wire bond called? Of what materials are the wires commonly used in wire bonding made of?

References

Akyildiz, I. F. and J. M. Jornet. 2010. Electromagnetic wireless nanosensor networks. *Nano Communication Networks* 1: 3–19.

Ambrosi, A., F. Airo, and A. Merkoc. 2010. Enhanced gold nanoparticle based ELISA for a breast cancer biomarker. *Analytical Chemistry* 82: 1151–1156.

Baik, N. S., G. Sakai, K. Shimanoe, N. Miura, and N. Yamazoe. 2000. Hydrothermal treatment of tin oxide sol solution for preparation of thin-film sensor with enhanced thermal stability and gas sensitivity. *Sensors and Actuators B: Chemical* 65(1–3): 97–100.

Biswas, K., S. Das, D. K. Maurya, S. Kal, and S. K. Lahiri. 2006. Bulk micromachining of silicon in TMAH-based etchants for aluminum passivation and smooth surface. *Microelectronics Journal* 37(4): 321–327.

Biswas, K. and S. Kal. 2006. Etch characteristics of KOH, TMAH and dual doped TMAH for bulk micromachining of silicon. *Microelectronics Journal* 37(6): 519–525.

Chen, Z. H., J. S. Jie, L. B. Luo, H. Wang, C. S. Lee1, and S. T. Lee. 2007. Applications of silicon nanowires functionalized with palladium nanoparticles in hydrogen sensors. *Nanotechnology* 18: 345502 (5pp), doi: 10.1088/0957-4484/18/34/345502

Childs, K., S. Dirk, S. Howell, R. J. Simonson, and D. Wheeler. 2005. SANDIA REPORT SAND2005–6004 Functionalized nanoparticles for sensor applications, 35pp.

Drbohlavova, J., V. Adam, R. Kizek, and J. Hubalek. 2009. Quantum dots—Characterization, preparation and usage in biological systems. *International Journal of Molecular Sciences* 10: 656–673, doi: 10.3390/ijms10020656

Hagerty, J., F. Helmbrecht, W. McCalpin, R. Zane, and Z. Popovic. 2004. Recycling ambient microwave energy with broad-band rectenna arrays. *IEEE Transactions on Microwave Theory and Techniques* 52(3): 1014–1024.

Haynes, C. L. and R. P. Van Duyne. 2001. Nanosphere lithography: A versatile nanofabrication tool for studies of size-dependent nanoparticle optics. *Journal of Physical Chemistry B* 105(24): 5599–5611, doi: 10.1021/jp010657m

Jamieson, T., R. Bakhshi, D. Petrova, R. Pocock, M. Imani, and A. M. Seifalian. 2007. Biological applications of quantum dots. *Biomaterials* 28: 4717–4732.

Jones, A. C. and M. L. Hitchman. 2009. Overview of chemical vapour deposition. In *Chemical Vapour Deposition: Precursors, Processes and Applications*. A. C. Jones and M. L. Hitchman (Eds.), Cambridge, U.K.: Royal Society of Chemistry, pp. 1–36.

Juttukonda, V., R. L. Paddock, J. E. Raymond, D. Denomme, A. E. Richardson, L. E. Slusher, and B. D. Fahlman. 2006. Facile synthesis of tin oxide nanoparticles stabilized by dendritic polymers. *Journal of American Chemical Society* 128(2): 420–421, doi: 10.1021/ja056902n

Kale, N. S., M. Joshi, P. N. Rao, S. Mukherji, and V. R. Rao. 2007. Bio-functionalization of silicon nitride-based piezoresistive microcantilevers. In *Proceedings of the 10th International Conference on Advanced Materials (ICAM)*, organized by the International Union of Material Research Societies (IUMRS), Bangalore, India, October 8–13, 2007.

Khanna, V. K. 2004. Emerging trends in ultra-miniaturized CMOS (complementary metal-oxide-semiconductor) transistors, single-electron and molecular-scale devices: A comparative analysis for high-performance computational nanoelectronics, *Journal of Scientific & Industrial Research* 63(10): 795–806.

Khanna, V. K. 2007a. Retrospection of manufacturing technologies and materials for deep submicron and nanometer CMOS integrated circuits, In *Proceedings of IMS-2007, Trends in VLSI and Embedded Systems*, J. N. Roy and D. Syal (Eds.), August 17–18, 2007, Punjab Engineering College, Chandigarh, pp. 19–27.

Khanna, V. K. 2007b. Advancing frontiers of MEMS and NEMS technologies: Micro- and nanomachining, In *Proceedings of IMS-2007, Trends in VLSI and Embedded Systems*, J. N. Roy and D. Syal (Eds.), August 17–18, 2007, Punjab Engineering College, Chandigarh, pp. 317–325.

Khanna, V. K. 2007c. Wet- and dry-etching based micromachining techniques in silicon MEMS fabrication, In *Proceedings of the All India Conference on Recent Developments in Manufacturing & Quality Management (RDMQM – 2007)*, Chief Editor P. B. Mahapatra, October 5th–6th, 2007, Punjab Engineering College, Chandigarh, pp. 46–52.

Li, J., Y. Lu, and M. Meyyappan. 2006. Nano chemical sensors with polymer-coated carbon nanotubes. *IEEE Sensors Journal* 6(5): 1047–1051.

MacKenzie, K. J., O. M. Dunens, C. H. See, and A. T. Harris. 2008. Large-scale carbon nanotube synthesis. *Recent Patents on Nanotechnology* 2: 25–40.

McFarland, A. D., C. L. Haynes, C. A. Mirkin, R. P. Van Duyne, and H. A. Godwin. 2004. Color my nanoworld. *Journal of Chemical Education* 81(4): 544A.

Mulfinger, L., S. D. Solomon, M. Bahadory, A. V. Jeyarajasingam, S. A. Rutkowsky, and C. Boritz. 2007. Synthesis and study of silver nanoparticles. *Journal of Chemical Education* 84(2): 322–325, doi: 10.1021/ed084p322

Murcia, M. J. and C. A. Naumann. 2005. Biofunctionalization of fluorescent nanoparticles. *Biofunctionalization of Nanomaterials*. Nanotechnologies for the Life Sciences. Vol. 1. Challa S. S. R. Kumar (Ed.), Weinheim, Germany: Wiley-VCH Verlag GmbH & Co., pp. 1–40.

Nayral, C., T. Ould-Ely, A. Maisonnat, B. Chaudret, P. Fau, L. Lescouzères, and A. Peyre-Lavig. 1999. A novel mechanism for the synthesis of Tin/Tin Oxide nanoparticles of low size dispersion and of nanostructured SnO^2 for the sensitive layers of gas sensors. *Advanced Materials* 11(1): 61–63, doi: 10.1002/(SICI)1521–4095(199901)

Pinna, N., G. Neri, M. Antonietti, and M. Niederberger. 2004. Nonaqueous synthesis of nanocrystalline semiconducting metal oxides for gas sensing. *Angewandte Chemie International Edition* 43(33): 4345–4349.

Pietryga, J. M., D. J. Werder, D. J. Williams, J. L. Casson, R. D. Schaller, V. I. Klimov, and J. A. Hollingsworth. 2008. Utilizing the lability of lead selenide to produce heterostructured nanocrystals with bright, stable infrared emission. *Journal of American Chemical Society* 130(14): 4879–4885, doi: 10.1021/ja710437r

Riegler, J., F. Ditengou, K. Palme, and T. Nann. 2008. Blue shift of CdSe/ZnS nanocrystal-labels upon DNA-hybridization. *Journal of Nanobiotechnology* 6: 7, doi: 10.1186/1477-3155-6-7

Rosso-Vasic, M., E. Spruijt, Z. Popovi, K. Overgaag, B. van Lagen, B. Grandidier, D. Vanmaekelbergh, D. Domínguez-Gutiérrez, L. De Cola, and H. Zuilhof. 2009. Amine-terminated silicon nanoparticles: Synthesis, optical properties and their use in bioimaging. *Journal of Materials Chemisry* 19: 5926–5933.

See, C. H. and A. T. Harris. 2007. A review of carbon nanotube synthesis via fluidized-bed chemical vapor deposition. *Industrial & Engineering Chemistry Research* 46(4): 997–1012, doi: 10.1021/ie060955b

Silva, R. S., A. F. G. Monte, P. C. Morais, A. M. Alcalde, F. Qu, and N. O. Dantas. 2006. Synthesis and characterization of PbS quantum dots embedded in oxide glass. *Brazilian Journal of Physics* 36(2A): 394–396.

Singh, P., R. M. Tripathi, and A. Saxena. 2010. Synthesis of carbon nanotubes and their biomedical application. *Journal of Optoelectronics and Biomedical Materials* 2(2): 91–98.

Tabatabaei, S., A. Shukohfar, R. Aghababazadeh, and A. Mirhabibi. 2006. Experimental study of the synthesis and characterization of silica nanoparticles via the sol–gel method. *Journal of Physics: Conference Series* 26: 371–374.

Tan, Y., X. Dai, Y. Li, and D. Zhu. 2003. Preparation of gold, platinum, palladium and silver nanoparticles by the reduction of their salts with a weak reductant–potassium bitartrate. *Journal of Materials Chemistry* 13: 1069–1075.

Tanaka, M. 2005. Spintronics: Recent progress and tomorrow's challenges. *Journal of Crystal Growth* 278(1–4): 25–37.

Vullum, F. and D. Teeters. 2005. Investigation of lithium battery nanoelectrode arrays and their component nanobatteries. *Journal of Power Sources* 146(1–2): 804–808.

Vullum, F., D. Teeters, A. Nytén, and J. Thomas. 2006. Characterization of lithium nanobatteries and lithium battery nanoelectrode arrays that benefit from nanostructure and molecular self-assembly. *Solid State Ionics* 177(26–32): 2833–2838.

Wang, Z. L. 2008. Towards self-powered nanosystems: From nanogenerators to nanopiezotronics. *Advanced Functional Materials* 18(22): 3553–3567.

Wang, Z., Q. Xu, H.-Q. Wang, Q. Yang, J.-H. Yu, and Y.-D. Zhao. 2009. Hydrogen peroxide biosensor based on direct electron transfer of horseradish peroxidase with vapor deposited quantum dots. *Sensors and Actuators B: Chemical* 138: 278–282.

Zhai, C., H. Zhang, N. Du, B. Chen, H. Huang, Y. Wu, and D. Yang. 2011. One-pot synthesis of biocompatible CdSe/CdS quantum dots and their applications as fluorescent biological labels. *Nanoscale Research Letters* 6: 31(5pp).

4

Mechanical Nanosensors

4.1 Introduction

Precise measurements of physical variables like mass, displacement, acceleration, force, torque, pressure, strain, fluid flow, viscoelasticity, and other properties of matter are required everyday in many applications. The main methods of sensing mechanical measurands have been established for many years and are therefore directly applicable to nanosensors. There is, however, a conspicuous effect that must be accounted for when considering micro- and nanoscale devices. This is, of course, *scaling*. Some physical effects favor the typical dimensions of smaller devices while others do not, e.g., as the linear dimensions of an object are reduced, other parameters do not cooperatively shrink in the same manner. Consider a simple cube of material of a given density. If the length l is reduced by a factor of 10, the volume (and hence mass) will be reduced by a factor of 1000 (l^3). Besides physical phenomena that can be extrapolated to nanoscale, many new phenomena that occur only at nanolevel, such as quantum-mechanical tunneling, single-electron transistor (SET) action, and so on, are exploited in the operation of mechanical nanosensors.

Scaling laws illustrate that on shrinking a body there is not only size reduction but the physical effects are also influenced (Wautelet 2001). Gravitational force between bodies dominates in the macroworld but at the microscopic level, adhesion forces are dominant, and these are primarily the van der Waals' forces between atoms and molecules. At the nanoscale, the gravitational force may be neglected in the analysis. As another example, the lowest frequency ν of a vibrating beam or string corresponds to a state where the length of the beam/string equals a quarter or half the wavelength λ. The well-known formula for phase velocity of the wave, $\upsilon = \nu\lambda$, means that resonance frequencies are larger in smaller systems. Citing the case of a cantilever beam, the deflection of a cantilever loaded by its own weight varies inversely as the square of its length so that a 10^3 times smaller beam bends 10^6 times more due to its own weight.

4.2 Nanogram Mass Sensing by Quartz Crystal Microbalance

The quartz crystal microbalance (QCM) is a mass-sensing device capable of measuring mass changes in the nanogram range (Kanazawa and Cho 2009). It is basically a quartz crystal resonator with the ability to measure very small mass changes in real time.

A *resonator* is a mechanical structure designed to resonate, that is, exhibit resonance or resonant behavior, naturally oscillating at some particular frequencies, called its *resonance frequencies*, with greater amplitude than at others. Resonators can be fabricated from a range of single crystal materials with micron-sized dimensions using various micromachining processes. The high dimensional stability and low temperature coefficient (the relative change of a physical property when the temperature is changed by 1 K) of quartz makes it a good resonator, keeping the resonant frequency constant.

The sensitivity of the QCM is approximately 100-fold that of an electronic balance (sensitivity ~ 0.1 mg). Elaborating further this comparison of QCM with other balances: depending on their maximal load, the balances used today in laboratories, e.g., beam, cantilever, torsion, or spring balances, can detect even 10^{-10} kg. On the other hand, the QCM can detect up to 10^{-16} kg. This has the implication that QCM is capable of measuring mass changes as small as a fraction of a monolayer or single layer of atoms.

What are the applications of QCM? Since the QCM was first introduced by Sauerbrey in 1959, it has become a largely used instrument for small mass measurements in vacuum, gas, and liquid phase. The high sensitivity and the real-time monitoring of mass changes on the sensor crystal render QCM a very attractive technique for a large range of applications.

What is the construction of QCM? The heart of the QCM is the piezoelectric AT-cut quartz crystal sandwiched between a pair of electrodes (Figure 4.1). Crystal blanks are cut at different orientations from the bar of quartz to realize specific desirable characteristics. The cut is determined by the specifications of the crystal. The keyhole-shaped electrodes on both major faces of the quartz resonator are vacuum-deposited gold or silver films, having a thickness of about 150 nm.

How does QCM work? When the electrodes are connected to an oscillator and an AC voltage is applied across these electrodes, the quartz crystal vibrates and thickness shear acoustic waves called bulk acoustic waves (BAWs) arising from the piezoelectric effect undergo constructive interference such that resonances occur at particular frequencies. Essentially, the quartz crystal starts oscillating at its resonance frequency. It is a piezoelectric resonator.

How mass is measured with QCM? If a rigid layer is evenly deposited on one or both the electrodes, the resonant frequency decreases proportionally to the mass of the adsorbed layer according to the Sauerbrey equation:
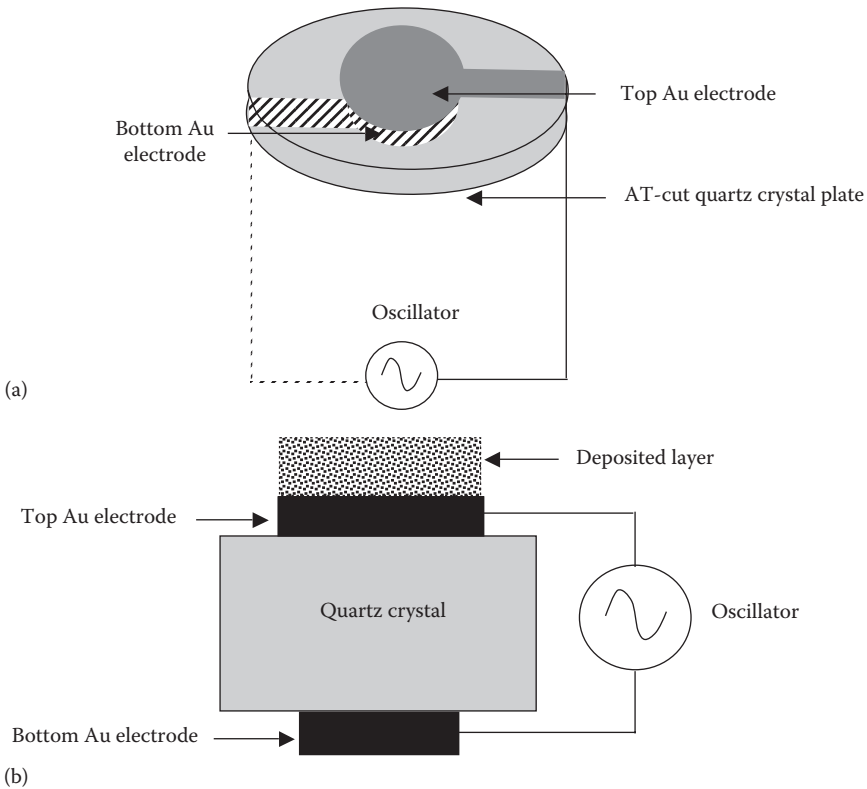
FIGURE 4.1
Quartz crystal microbalance: (a) schematic diagram and (b) cross section.

$$\Delta f = -\frac{2f_{\mathrm{o}}^{2}\Delta m}{A\sqrt{\rho_{\mathrm{q}}\mu_{\mathrm{q}}}} \qquad (4.1)$$

where
  Δf is the measured frequency shift
  f_{o} is the resonant frequency of the fundamental mode of the crystal
  Δm is the mass change per unit area (g cm^{-2})
  A is the piezoelectrically active area of the device
  ρ_{q} is the density of quartz, 2.648 g cm^{-3}
  μ_{q} is the shear modulus of quartz, 2.947×10^{11} g cm$^{-1} \times$ s^{-2}

The resonant frequency changes as a linear function of the mass of material deposited on the crystal surface. The mass-sensitive area is situated in the

central part of the resonator, covering about the area where the two electrodes overlap.

Under what conditions does Sauerbrey equation apply? The Sauerbrey equation is derived by treating the deposited mass as though it were an extension of the thickness of the underlying quartz. Hence, it is only strictly applicable to uniform, rigid, thin-film deposits. In certain situations, the Sauerbrey equation does not hold; citable instances are when the added mass is (i) not rigidly deposited on the electrode surface(s); (ii) slips on the surface; or (iii) is not deposited evenly on the electrode(s).

On the contradictory side, because of the method of derivation of Sauerbrey's equation, the mass to frequency correlation, as determined by this equation, is largely independent of electrode geometry. This has the benefit of allowing mass determination without calibration, making the setup desirable from a cost and time investment standpoint.

Sauerbrey also developed a method for measuring the characteristic frequency and its changes by using the crystal as the frequency-determining component of an oscillator circuit. His method still remains the primary tool in QCM experiments for conversion of frequency to mass and is valid in nearly all applications.

QCM is a BAW device. For comparison, a surface acoustic wave (SAW) gas sensor is shown in Figure 4.2 (Ho et al. 2003). It consists of the gas-sensitive material deposited between the input and output interdigitated transducers (IDTs). The input IDT creates electric fields alternating in time and space, which penetrate into piezoelectric substrate and generate mechanical stresses through the piezoeffect. In this way, the transducer produces SAWs, which propagate on the surface. These waves are detected by the output transducer.

The gas gets adsorbed on the sensitive film when the SAW sensor is exposed to it at room temperature and/or at elevated temperature, which in turn changes conductivity of the film. This conductivity change affects the velocity of the SAW traveling across the film and gives rise to a frequency change, which corresponds to the percentage of the gas molecules adsorbed by the film.

A *delay line* is a circuit designed to introduce a calculated delay into the transmission of a signal. An *acoustic delay line* delays the propagation of a sound wave by circulating it through a liquid or solid medium. *SAW delay lines* are used as sensitive mass detectors by virtue of changes in SAW velocity.

Decreasing the SAW device area increases the operating frequency, f, and dramatically improves the mass sensitivity (the mass detection limit is proportional to $1/f^3$, assuming that noise is linear in f). The operating frequency of SAW devices is inversely proportional to the spatial periodicity of the transducers, so that central frequencies in the GHz range for sensor require submicron or nanometric lithography.
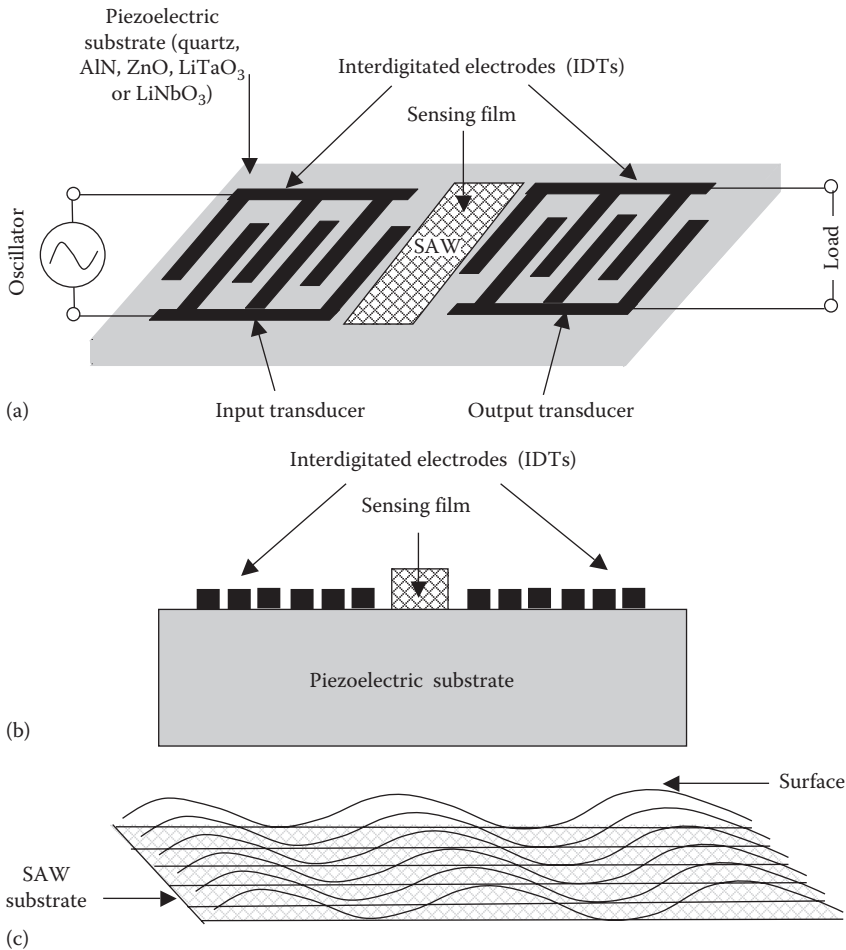
FIGURE 4.2
Surface acoustic wave (SAW) device: (a) schematic layout, (b) cross-sectional diagram, and (c) SAWs.

4.3 Attogram (10^{-18} g) and Zeptogram (10^{-21} g) Mass Sensing by MEMS/NEMS Resonators

Sensor designs based on QCM have been refined over decades and are approaching their theoretical limitations. Some of the drawbacks of quartz crystals are (i) their relatively large sizes along with their (ii) sensitivity to shock and vibration, and (iii) their higher failure rates than silicon

components. For ultrasensitive mass detection, silicon micro-electrome-chanical system (MEMS)-based micro-resonator is better than QCM sensor because of (i) its advantage of having a much smaller minimum detectable mass limit by several orders of magnitude. (ii) Apart from this, MEMS-based micro-resonators exhibit better flexibility in device design and fabrication for various kinds of specimens, and (iii) offer better integration capability with other MEMS components and complementary metal-oxide semicon-ductor (CMOS) circuits.

Nanoelectromechanical systems (NEMS) have provided the unsurpassed opportunity for gaining insight into physical phenomena and chemical interactions (Ekinci et al. 2004, Dai et al. 2009). The fundamental principle in nanomechanical detection using NEMS resonator is the direct transduction of molecular adsorption into the resonant frequency shift arising from atomic mass of adsorbed molecules (Sone et al. 2004). Reducing the size of a resonator increases its resonant frequency. Since frequencies in the GHZ range are pos-sible, NEMS resonators are envisioned for RF or microwave communication.

Why should the mass of the resonator itself be small? A consequence of the smaller mass of the resonator is that smaller changes in the mass are detect-able, which confers on a NEMS resonator the capability for perception of masses approaching the level of single atoms. Thus smaller is mechanical resonator, the better sensitivity does it exhibit. Based on such a concept, NEMS resonators have been suggested as nanosensors that allow the sen-sitive detection of molecules at atomic resolution. This is one field where nanotechnology has an unquestionable position.

Various types of resonator structures include cantilevered resonators (Li et al. 2003, Jenkins et al. 2004, Lin et al. 2005, Hansen and Thundat 2005, Ikehara et al. 2007, Rinaldi 2009), clamped–clamped beam, free–free beam (Figures 4.3 and 4.4), etc. (Beeby et al. 2004).
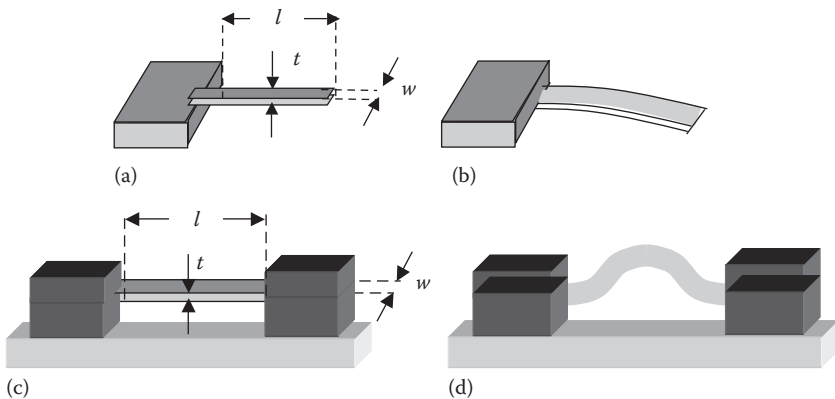
FIGURE 4.3
(a) Straight cantilever, (b) bent cantilever, (c) straight doubly clamped beam resonator, and (d) vibrating doubly clamped beam. Symbols: length *l*, width *w*, and thickness *t*.
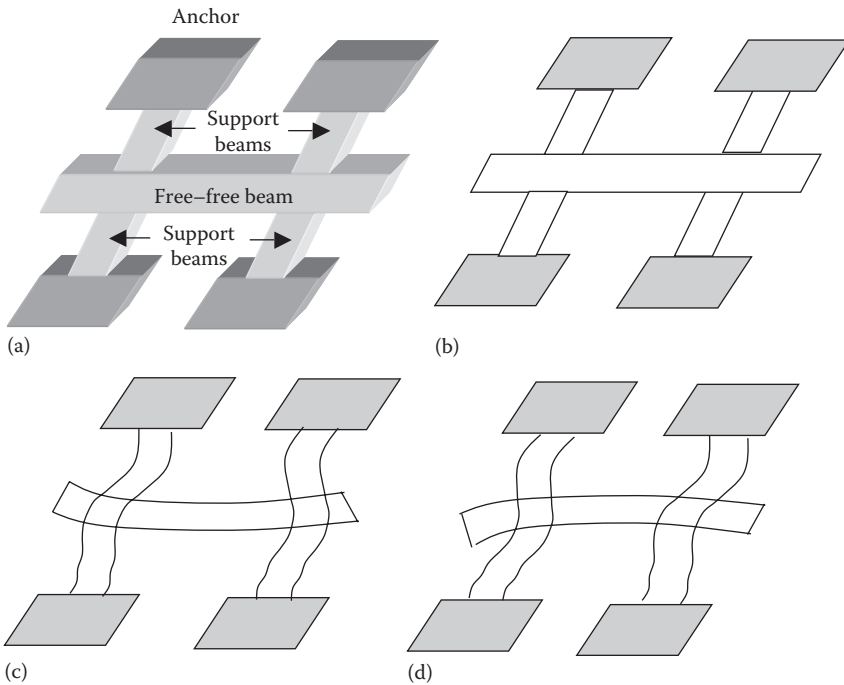
FIGURE 4.4
Free–free beam: (a) and (b) 3D and 2D views in equilibrium; (c) and (d) vibrating positions.

4.3.1 Microcantilever Definitions and Theory

A *cantilever* is a projecting structure, such as a beam, that is supported rigidly at one end and carries a load at the other end or along its length. Cantilevers are fabricated in several shapes (Yang et al. 2000, Ansari et al. 2009): rectangular, two-legged, and V-shaped (Figure 4.5). The V-shape is the predominant cantilever manufactured and used currently in AFM. The popular V-shaped cantilever is intuitively thought to be resistant to lateral forces and therefore less prone to twisting than rectangular beam. But V-shaped cantilever makes calibration of the microscope and interpretation of data more difficult. The two-leg type cantilever reduces the spring constant. By decreasing the spring constant, a significant performance improvement is achieved, through the corresponding increase in sensitivity.

   A microcantilever is the miniaturized counterpart of a diving board that moves up and down at a regular interval undergoing a vibrational motion (Figure 4.6a). It is a miniature diving board anchored at one end to a relatively large mass (Vashist 2007). The movement of the microcantilever changes when a specific mass of analyte is specifically adsorbed on its surface similar to the change when a person steps onto the diving board. Molecules adsorbed on a microcantilever cause deflection (Figure 4.6b) or vibrational
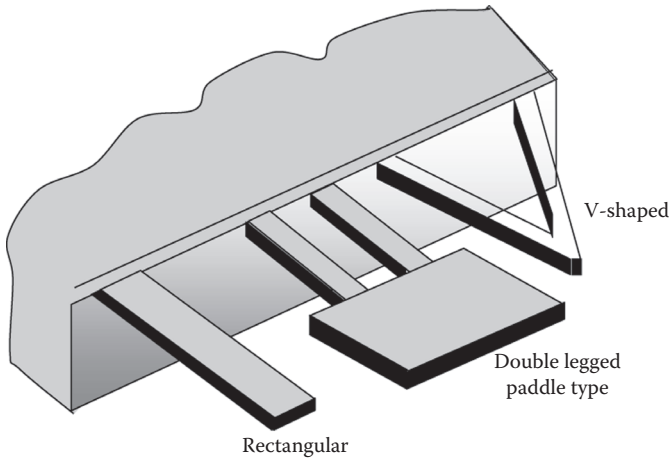
FIGURE 4.5
Cantilevers of different shapes (Vashist 2007).
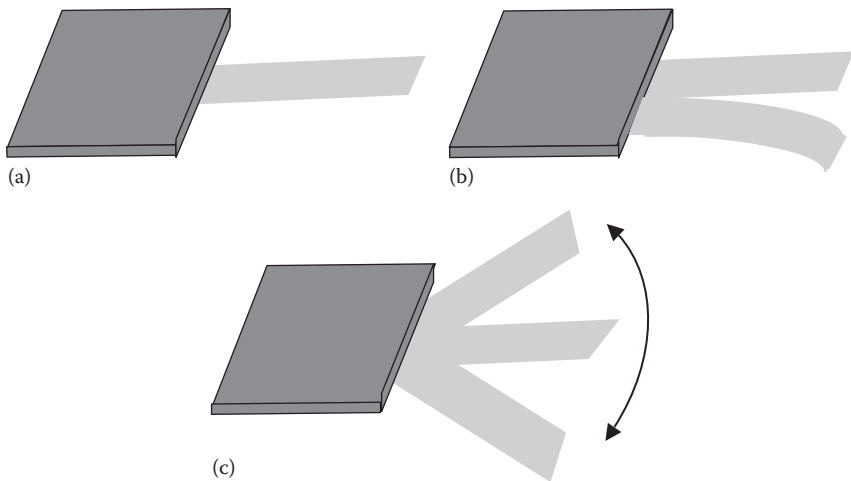
FIGURE 4.6
Cantilevers: (a) straight cantilever, (b) bent cantilever, and (c) vibrating cantilever.

frequency changes of the microcantilever (Figure 4.6c). Many sensor applications require arrays comprising several cantilevers each of which is assigned a specified task (Figure 4.7).

What are the operational modes of a cantilever? Modes of cantilever operation can be divided into static and dynamic modes from the viewpoint of the measured parameter (Lavrik et al. 2004). In static mode, the static bending or deflection of the cantilever is measured while in dynamic mode its dynamic
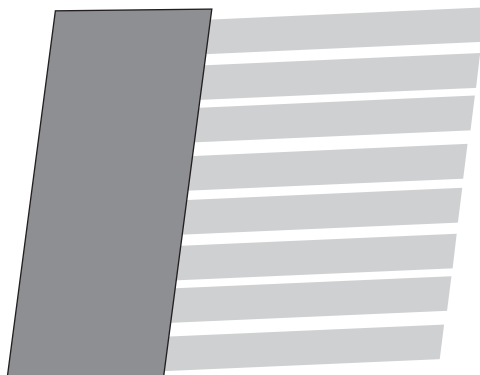
FIGURE 4.7
An array of cantilevers.

properties like resonance frequency are determined. Each of these modes, in turn, is associated with different transduction scenarios. Static cantilever deflections are caused by one of the two primary causes: either external forces exerted on the cantilever as in AFM or intrinsic stresses generated on the cantilever surface or within the cantilever. While cantilever microfabrication technology is capable of producing nearly stress-free suspended beams, additional intrinsic stresses may subsequently be produced from several factors like thermal expansion, interfacial processes, and physicochemical changes.

How do viscous properties of the medium influence cantilevers in their two operating modes? Cantilever sensors operating in the dynamic mode are essentially mechanical oscillators, resonance characteristics of which depend upon both the attached mass and viscoelastic properties of the medium, e.g., adsorption of analyte molecules on a resonating cantilever results in lowering of its resonance frequency due to the increased suspended mass of the resonator (Chen et al. 1995). The damping effects (effects that tend to reduce the amplitude of oscillations) of a liquid medium reduce resonance responses of cantilever devices. In most common liquids, such as aqueous solutions, the amplitude of the cantilever oscillations at the resonance is orders of magnitude lower as compared to the same resonating cantilever operating in air. On the other hand, operation in the static mode is unaffected by viscous properties of the medium. Therefore, microcantilever sensors operating in the static mode are especially attractive as platforms for nanomechanical biochemical assays and other biomedical applications.

How is the small deflection of a cantilever, often in nm, measured? There are two methods: (i) optical and (ii) electrical (Figure 4.8). In the *optical method*, shown in Figure 4.8a, a laser beam reflected from a gold film–coated spot on the surface of the cantilever is received by a photodetector, both in the unbent and bent states of the cantilever. The deflection is determined from the
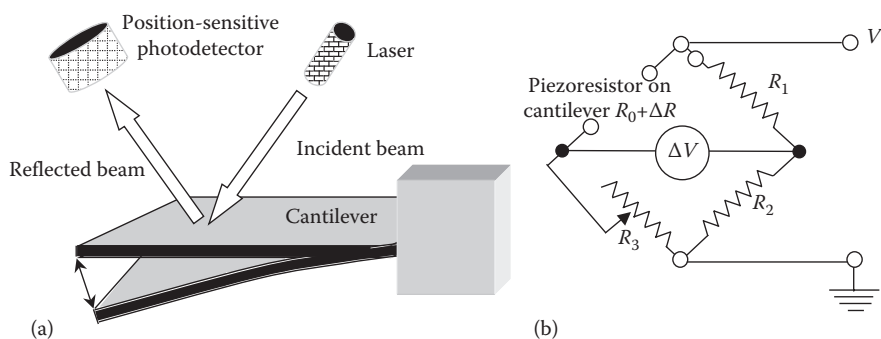
FIGURE 4.8
Measurement of cantilever deflection by: (a) optical deflection system and (b) Wheatstone bridge circuit in which the piezoresistor on cantilever located near the end connected to its support is put in one arm of the bridge.

difference in position of the reflected beam for the two cases, i.e., for the bent beam with respect to unbent beam. The method provides sub-nanometer accuracy. Another optical method is based on interferometry (investigative technique involving the design and use of optical inferometers) utilizing the interference between a reference laser beam and the one reflected from the cantilever but is less suitable for liquids and hence finds limited usage in biosensing. Besides the aforementioned methods, optical diffraction grating and charged coupled device are also used for deflection measurements. The *diffraction grating* is a component with a periodic structure, which splits and diffracts light into several beams traveling in different directions. The *charge-coupled device* (*CCD*) is a light-sensitive integrated circuit consisting of an array of closely spaced MOS capacitors that can store and transfer information using packets of electric charge. Each capacitor in the array represents a pixel so that by applying an external voltage to the top plates of the MOS structure, charges (electrons or holes) can be stored in the resulting potential well. These charges can be shifted from one pixel to another by digital pulses applied to the top plates (gates), and in this way the charges can be transferred row by row to a serial output register. The picture represents the display of the electron distribution.

In the *electrical method*, illustrated in Figure 4.8b, a polysilicon piezoresistor is formed on the cantilever surface at the position of maximum stress near the supporting point of the cantilever and the deflection is obtained from the change in resistance value of this resistor as determined by the upsetting of the balance of the Wheatstone bridge; the *Wheatstone bridge* is a circuit containing four resistances, a constant voltage input, and a voltage gage, that is used to determine the value of an unknown resistance when the other three resistances are known. The alternative capacitance method is unsuitable for large displacements and also for electrolytic solutions.

How is a cantilever beam actuated for oscillations and how is its resonance frequency measured?

1. In *magnetomotive motion actuation*, the Lorentz force generated upon a current-carrying conductor in a static magnetic field supplies the actuation force. Here, an AC current is driven through the beam in the presence of a strong magnetic field.

2. In *electrostatic (capacitive) motion actuation*, a gate electrode is fabricated in the vicinity of the beam and a voltage is applied between the gate electrode and the device. The actuation force depends nonlinearly upon the excitation voltage and has components present at higher harmonics. To obtain a large contribution at the drive frequency, a voltage having a large DC and a small AC component is usually applied.

3. An important actuation technique is based on piezoelectricity. A piezoelectric material responds to an externally applied electric field by deforming. Among various types of micro-resonators, piezoelectric thin-film transduced micro-resonator offers competitive applications because of its unique characteristics of low power consumption, small driving voltage, self-actuation self-sensing capability, and impedance matching with CMOS circuits.

How do these cantilevers interface well with CMOS circuitry? Piezoelectric micro-resonators provide better power handing capability than capacitive resonators since low driving voltage of several 100 mV is adequate for resonator actuation, and this facilitates the integration of micro-resonators with CMOS circuits. Lu et al. (2007) fabricated a partially PZT-covered single-crystal silicon cantilever for ultrasensitive mass detection applications. PZT stands for lead zirconate titanate [PZT, $Pb(Zr_{0.52},Ti_{0.48})O_3$] film. The PZT film was integrated only at the fixed end of the cantilever to suppress energy dissipation and other negative effects as well as to improve the mechanical properties of the cantilever. The 30 μm-wide 100 μm-long cantilever (Figure 4.9) was fabricated on silicon-on-insulator wafers. The PZT was used as both the microactuator to drive the microcantilever and as a microsensor to determine its resonant frequency.

How is vibration analysis of a cantilever (or in general, any mechanical microstructure) performed? Laser Doppler Vibrometer (LDV) is a state-of-the-art facility for vibration analysis of a wide range of mechanical systems (Figure 4.10) (Yang et al. 2000). LDV is a noncontact technique for velocity and displacement measurement of vibrating structures so that the tested device is uninfluenced by the measurement process giving a reliable assessment of its properties. The laser beam from the LDV is directed at the object of interest, and its vibration amplitude and frequency are extracted from the Doppler shift of the laser beam frequency due to the motion of the object. *Doppler effect* describes the
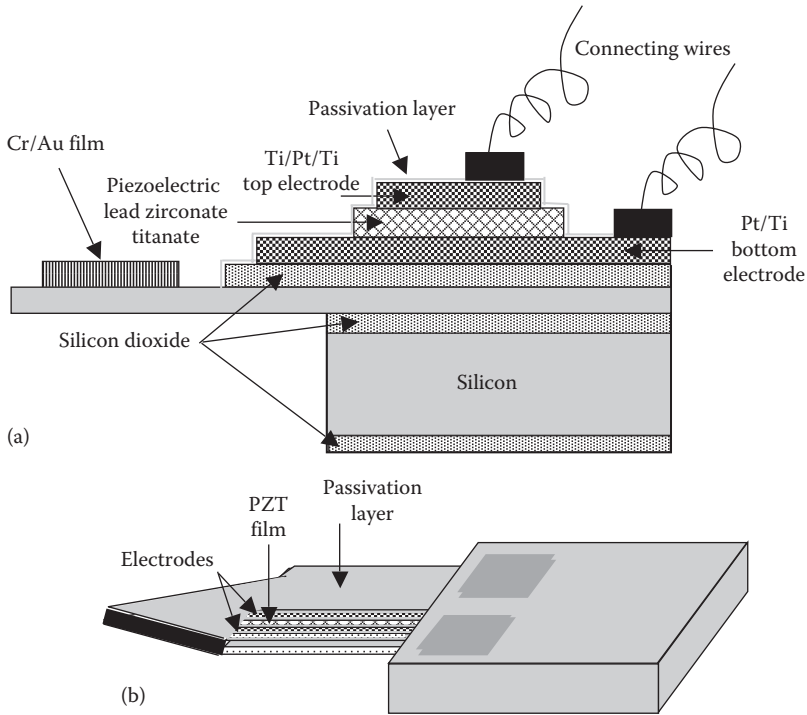
FIGURE 4.9
PZT-transduced cantilever for mass sensing: (a) cross section and (b) top and side view. (After Lu, J. et al., *Jpn. J. Appl. Phys.,* 46, 7643, 2007.)

relative change in wavelength and frequency of a wave when the source and observer are in relative motion with respect to each other.

A vibrometer is generally a two-beam laser interferometer that measures the frequency (or phase) difference between an internal reference beam and a test beam. A coherent (waves having a constant difference in phase) laser beam is divided into object and reference beams by a beam splitter 1. The object beam strikes a point on the moving or vibrating object and light reflected from that point travels back to beam splitter 2 and mixes or interferes with the reference beam at beam splitter 3. If the object is moving or vibrating, this mixing process produces an intensity fluctuation in the light. Whenever the object has moved by half the wavelength, $\lambda/2$, which is $0.3169\,\mu m$ in the case of He–Ne laser, the intensity goes through a complete dark–bright–dark cycle. A detector converts this signal to a voltage fluctuation. The Doppler frequency f_D of this sinusoidal cycle is proportional to the velocity υ of the object. Instead of detecting the Doppler frequency, the velocity is directly obtained by a *digital quadrature demodulation method* which involves translating a given bandpass signal to baseband signal and producing from it the in-phase and quadrature components. In the quadrature
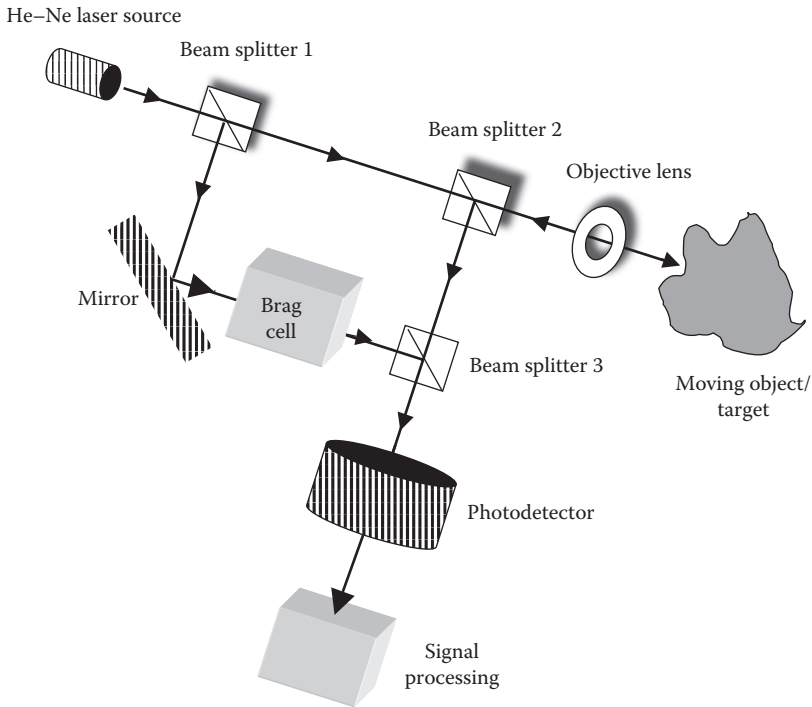
FIGURE 4.10
Modules of a laser Doppler vibrometer (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.
1.68.417&rep=rep1&type=pdf).

demodulator, the modulated carrier is passed through an inductance–capacitance (LC) tank circuit that shifts the signal by 90° at the center frequency. This phase shift is either greater or less than 90° depending on the direction of deviation. A phase detector compares the phase-shifted signal to the original to give the demodulated baseband signal.

The Bragg cell, which is an acousto-optic modulator to shift the light frequency by 40 MHz, is used for identifying the sign of the velocity.

What are the advantages of an LDV over similar measurement devices such as an accelerometer? The LDV can be directed at targets that are difficult to access, or that may be too small or too hot to attach a physical transducer. Also, the LDV makes the vibration measurement without mass-loading the target, which is especially important for MEMS devices.

Returning to the discussion on cantilevers, what benefits accrue with using cantilevers? One great advantage of the cantilever technique is that its five response parameters (resonance frequency, phase, amplitude, Q-factor, and deflection) can be simultaneously measured. A compelling feature of the cantilever-based sensors operating in the resonant mode is that four response parameters (resonance frequency, phase, amplitude, and Q-factor, measured

simultaneously) may provide complementary information about the interactions between the sensor and the environment. The bending and resonance frequency shifts of a cantilever are measurable with sub-Angstrom resolution using current techniques perfected for the AFM, such as optical beam deflection, piezoelectric, piezoresistive, and capacitance measurement methods.

Now describing the theory of microcantilever, simple beam theory is restricted to a prismatic (equal cross section), homogeneous, straight, and untwisted structure (Figure 4.11). The beam is a structural member, usually horizontal, whose main function is to carry loads transverse to its longitudinal
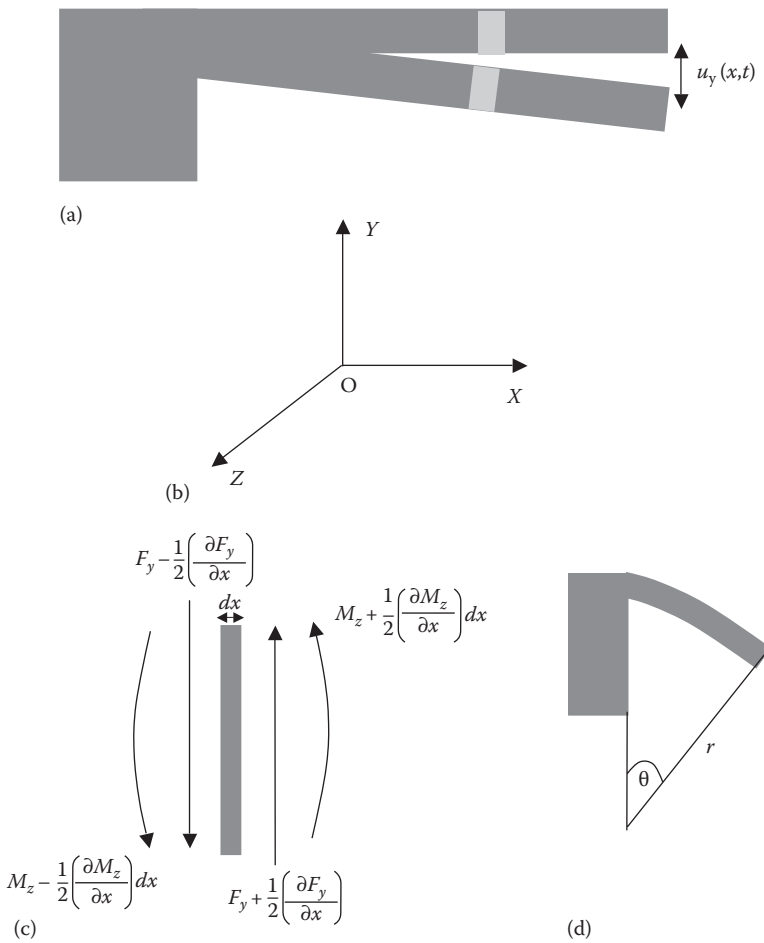
FIGURE 4.11

Flexural behavior of a straight beam: (a) prismatic beam, (b) co-ordinate system, (c) force and bending moment components, and (d) strain in the cantilever. (After Duemling, M., *Modeling and Characterization of Nanoelectromechanical Systems*. MS Thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA, 6, 2002.)

axis. These loads usually cause bending of the beam member. The beam is defined as a structure having two of its dimensions much smaller than the third. The thickness (t) and width (w) of the cantilever are small compared to the length (L), which reduces the analysis to a one-dimensional problem along the length of the beam. Additionally, it is presumed that the normal stresses (σ) in the y and z direction are negligible. The following derivation only holds if the maximum deflection (in y) is smaller than the radius of gyration (K). The *radius of gyration* of an object describes its dimensions. The radius of gyration of an object or body about a given axis is computed in terms of the moment of inertia around its center of gravity or a specified axis, and the total mass. It equals the square root of the ratio of the moment of inertia of the body about the given axis to its mass. Moment of inertia is the name given to rotational inertia, the rotational analogue of mass for linear motion. It plays the same role in rotational dynamics as mass in linear dynamics. The measure of the inertia in the linear motion is the mass of the system and its angular counterpart is the so-called moment of inertia. The moment of inertia of a body is not only related to its mass but also the distribution of the mass throughout the body. It must be specified with respect to a chosen axis of rotation. For a point mass, the moment of inertia is simply the mass times the square of perpendicular distance to the rotation axis.

Because K represents the root mean square distance of the infinitesimal parts constituting the object from the axis, if all the mass of the body is considered to be placed a distance K from the axis, the moment if inertia would be the same. Hence, it is a measure of how far the mass of the body is concentrated from its center of mass.

In beam theory, K is the distance from the neutral axis of a section to an imaginary point at which the whole area of the section could be concentrated and still have the same moment of inertia. It equals the square root of the moment of inertia divided by the area of the section.

If the maximum deflection approaches K, nonlinear terms have to be considered. The coordinate system used for the following derivation is shown in Figure 4.6b.

4.3.1.1 Resonance Frequency Formula

In Figure 4.11, the beam tip is under a uniformly distributed loading, σ. Its unit is Newton meter^{-1}, not Pascal and $\sigma > 0$ is tensile. σ is called the surface stress. This surface stress effect is modeled as a concentrated moment applied at the free end of the cantilever beam (Duemling 2002).

The problem is simplified if all shear (force tending to cause deformation of a material by slippage along a plane parallel or tangential to the imposed stress) and rotational forces are insignificant. The formulation is called *Euler–Bernoulli beam theory*, also known as engineer's beam theory or classical beam theory. It covers the case for small deflections of a beam which is subjected to lateral loads only.

The only remaining normal stress σ_x can be written as

$$\sigma_x = ky \tag{4.2}$$

where
  k is a constant
  $y=0$ lies in the center of the beam

The total internal force has to be zero, and it is given by

$$F_{int} = \int_A \sigma_x dA = 0 \tag{4.3}$$

A bending moment is a term used to describe the force or torque exerted on a material and leads to the event of bending or flexure within that material. Moments, internal or external, are rotation equivalents of forces, equal to force times distance. The bending moment in a beam is the moment, internal to the beam, necessary to counteract externally applied moments. The bending moment at a section through a structural element is defined as the sum of the moments about that section of all external forces acting to one side of that section. Without application of any external momentum, the total bending moment = the moment due to internal forces, which are only nonvanishing in the z direction:

$$M = M_z = \int_A y\sigma_x dA = k\int_A y^2 dA \tag{4.4}$$

Area moment of inertia, also known as *second moment of inertia*, is a property of shape that is used to predict deflections and stresses in beams. The area moment of inertia of a cross-sectional area of a beam measures ability of the beam to resist bending. The larger the moment of inertia, the less the beam will bend. This moment of inertia is a geometrical property of a beam and depends on a reference axis. The smallest moment of inertia about any axis passes through the *centroid*. The following is the mathematical equation to calculate the second moment of inertia:

$$I_z = \int_A y^2 dA \tag{4.5}$$

where
  I_z is the second moment of inertia
  y is the perpendicular distance from axis z to the element dA
  dA is an elemental area

k is defined by substituting Equation 4.5 into Equation 4.4 as

$$k = \frac{M_z}{I_z} \qquad (4.6)$$

The stress of the cross section is therefore given by

$$\sigma_x = \frac{M_z y}{I_z} \qquad (4.7)$$

Using Hook's law (for relatively small deformations of an object, the displacement or size of the deformation is directly proportional to the deforming force or load), the strain (ε) is calculated as

$$\varepsilon_x = \frac{\sigma_x}{E} \qquad (4.8)$$

where E is Young's modulus (longitudinal stress/longitudinal strain).

If $u_y(x, t)$ is the displacement of the beam in y direction and the deflection is small ($du_y/dy \ll 1$), then the second derivative of the deflection is $\approx$ the inverse of the radius of curvature r

$$\frac{\partial^2 u_y(x,t)}{\partial x^2} = \frac{1}{r} \qquad (4.9)$$

The strain can be found as

$$\varepsilon = \frac{dL - dL_o}{dL_o} = \frac{(r-y)d\theta - rd\theta}{r \sin d\theta} = -\frac{y}{r} \qquad (4.10)$$

Combining Equations 4.8 through 4.10 gives the *Euler–Bernoulli law* of elementary beam theory

$$M_z = -EI_z \left\{ \frac{\partial^2 u_y(x,t)}{\partial x^2} \right\} \qquad (4.11)$$

In absence of any external forces or bending moments acting on the beam, the equation of motion is

$$m \left\{ \frac{\partial^2 u_y(x,t)}{\partial t^2} \right\} = \sum F_{int} \qquad (4.12)$$

and the total momentum has to be zero

$$\sum M_{int} = 0 \tag{4.13}$$

The total force is calculated

$$\sum F_{int} = \left\{ F_y + \frac{1}{2}\left(\frac{\partial F_y}{\partial x}\right)dx \right\} - \left\{ F_y - \frac{1}{2}\left(\frac{\partial F_y}{\partial x}\right)dx \right\} = \left(\frac{\partial F_y}{\partial x}\right)dx \tag{4.14}$$

and the sum of the bending moments is

$$\sum M_{int} = \left\{ M_z + \frac{1}{2}\left(\frac{\partial M_z}{\partial x}\right)dx \right\} - \left\{ M_z - \frac{1}{2}\left(\frac{\partial M_z}{\partial x}\right)dx \right\}$$

$$+ \left\{ F_y + \frac{1}{2}\left(\frac{\partial F_y}{\partial x}\right)dx \right\}\frac{dx}{2} - \left\{ F_y - \frac{1}{2}\left(\frac{\partial F_y}{\partial x}\right)dx \right\}\left(-\frac{dx}{2}\right) \tag{4.15}$$

Combination of Equations 4.13 and 4.15 give the relationship interlinking the bending moment and the force

$$F_y = -\frac{\partial M_z}{\partial x} \tag{4.16}$$

Using Equations 4.16, 4.14, and the mass

$$m = \rho A dx \tag{4.17}$$

where
  ρ is the density
  A is the cross section

the equation of motion is expressed as

$$\rho A \frac{d^2 u_y(x,t)}{dt^2} + EI_z \frac{\partial^4 u_y(x,t)}{\partial x^4} = 0 \tag{4.18}$$

This harmonic linear fourth order differential equation is solved using *separation of variables* method. For a partial differential equation in a function $\Phi(x, y, \ldots)$ and variables $x, y, \ldots$, separation of variables can be applied by making a substitution of the form $\Phi(x, y, \ldots) \equiv X(x)Y(y), \ldots$, breaking the

resulting equation into a set of independent ordinary differential equations, solving these for $X(x)$, $Y(y)$, …, and then plugging them back into the original equation. This technique works because if the product of functions of independent variables is a constant, each function must separately be a constant.

Here, the complete solution is not of interest to us, but only the natural resonant frequency of the beam, which is easily obtained by using a Fourier transformation. The Fourier transform, in essence, decomposes or separates a function into sinusoids of different frequencies which sum to the original waveform. The Fourier transform is based on the fact that it is possible to take any periodic function and resolve it into an equivalent infinite summation of sine waves and cosine waves with frequencies that start at 0 and increase in integer multiples of a base frequency.

Applying a Fourier transformation to Equation 4.18 with

$$\phi\{u_y(x,t)\} = U_y(x,\omega) \tag{4.19}$$

yields

$$\rho A(i\omega)^2 U_y(x,\omega) + EI_z \frac{\partial^4 U_y(x,\omega)}{\partial x^4} = 0 \tag{4.20}$$

The calculation becomes simpler if the equation is rewritten as

$$-\alpha^4 \omega_y^2 U_y(x,\omega) + \frac{\partial^4 U_y(x,\omega)}{\partial x^4} = 0 \tag{4.21}$$

with

$$\alpha = \sqrt[4]{\frac{\rho A}{EI_z}} \tag{4.22}$$

The solution of this differential equation is

$$U_y(x,\omega) = A_1 \exp\left(\alpha x \sqrt{\omega}\right) + A_2 \exp\left(-\alpha x \sqrt{\omega}\right) + A_3 \exp\left(i\alpha x \sqrt{\omega}\right) + A_4 \exp\left(-i\alpha x \sqrt{\omega}\right) \tag{4.23}$$

where A_1, A_2, A_3, and A_4 are complex constants which are determined from the boundary condition. Using the Euler equations (the relationship between the trigonometric functions and the complex exponential function, Euler's formula for the complex exponential states that, for any real number θ, $e^{i\theta} = \cos\theta + i\sin\theta$ where e is the base of the natural logarithm, i is the

imaginary unit, and cos and sin are the trigonometric functions cosine and sine) and the comparable equations for sinh and cosh [sinh $x = (e^x - e^{-x})/2$ and cosh $x = (e^x + e^{-x})/2$], Equation 4.23 can be transformed to an equation with real constants B_1, B_2, B_3, B_4 as

$$U_y(x,\omega) = B_1 \sin\left(\alpha x\sqrt{\omega}\right) + B_2 \cos\left(\alpha x\sqrt{\omega}\right) + B_3 \sinh\left(\alpha x\sqrt{\omega}\right) + B_4 \cosh\left(\alpha x\sqrt{\omega}\right)$$

(4.24)

By applying the boundary conditions of the problem to Equation 4.24, the resonant frequency is found. In the following discussion, a cantilever will be considered. Let us focus our attention on the clamped side and free end of the beam. Since at the clamped side of the cantilever ($x=0$), no displacement takes place and the beam is straight, the boundary conditions are given by

$$U_y(0,\omega), \quad \frac{dU_y(0,\omega)}{dx} = 0$$

(4.25)

At the free end of the beam ($y=L$), there is no bending moment and no shear forces act on the beam:

$$\frac{d^2U_y(l,\omega)}{dx^2} = 0, \quad \frac{d^3U_y(l,\omega)}{dx^3} = 0$$

(4.26)

Applying the first two boundary conditions, we have $B_2 = B_4$ and $B_1 = -B_3$. The last two boundary conditions reduce Equation 4.24 to

$$\frac{2 + 2\cos\left(\alpha l\sqrt{\omega}\right)\cosh\left(\alpha l\sqrt{\omega}\right)}{\sin\left(\alpha l\sqrt{\omega}\right) - \sinh\left(\alpha l\sqrt{\omega}\right)} = 0$$

(4.27)

A nontrivial solution (a solution of a set of equations in which at least one of the variables has a value different from zero) is found if

$$\cos\left(\alpha l\sqrt{\omega}\right)\cosh\left(\alpha l\sqrt{\omega}\right) = -1$$

(4.28)

There is no analytical solution (one that can be explicitly written down, also called a closed form solution) for this equation but it is solved numerically by using the substitution

$$\beta = \alpha l\sqrt{\omega}$$

(4.29)

TABLE 4.1

β_i^2 Values for Different Beams

Type of Beam	Value of b_i^2 for $i =$				
	1	2	3	4	5
Cantilever	3.516	22.034	61.701	120.912	199.855
Clamped–clamped beam	22.373	61.678	120.903	199.860	298.526
Free–free beam	22.373	61.678	120.903	199.860	298.526

The values for β_i are given in Table 4.1. From Equations 4.22 and 4.21, the natural resonant frequency and its harmonics (frequencies that are integral multiples of the fundamental frequency) are calculated:

$$\omega_i = \left(\frac{\beta_i^2}{l^2}\right)\sqrt{\frac{EI_z}{\rho A}} \tag{4.30}$$

For $i = 1$, $\beta_i^2 = 3.516$; and for a rectangular cantilever

$$I_z = \frac{wt^3}{12} \tag{4.31}$$

Hence from Equation 4.30, we get

$$\omega_1 = \left(\frac{3.516}{l^2}\right)\sqrt{\frac{Ewt^3}{12\rho tw}} = 1.015\sqrt{\frac{E}{\rho}}\left(\frac{t}{l^2}\right) \tag{4.32}$$

The fundamental frequency ω_1 is a function of the material parameters E and ρ as well as the beam dimensions t and l. High frequencies are achieved by reducing the overall cantilever scale, by choosing stiffer (not easily bent or changed in shape) and lighter materials, and by reducing the aspect ratio l/t.

Example 4.1

Two types of cantilevers (200 µm [length] × 100 µm [width] × 10 µm [thickness]; 6 µm [length] × 1 µm [width] × 0.17 µm [thickness]; 10 µm [length] × 0.6 µm [width] × 0.75 µm [thickness]) are fabricated from silicon, silicon nitride, and silicon carbide. The properties of these materials are given in Table 4.2. Calculate the angular resonant frequencies of these cantilevers.

For cantilever length = 200 µm, width = 100 µm, thickness = 10 µm, from Equation 4.32,

TABLE 4.2

Young's Modulus and Density of Materials (See Example 4.1)

Material	Young's Modulus (Y) in N m^{-2}	Density (ρ) in kg m^{-3}
Silicon	1.7×10^{11}	2329
Silicon nitride	3.1×10^{11}	3290
Silicon carbide	4.1×10^{11}	3100

$$\omega_1 = 1.015 \sqrt{\frac{E}{\rho}} \left(\frac{t}{l^2}\right) = 1.015 \left\{\frac{10 \times 10^{-6}}{\left(200 \times 10^{-6}\right)^2}\right\} \sqrt{\frac{E}{\rho}} = 250 \sqrt{\frac{E}{\rho}} \qquad (4.33)$$

For cantilever length $= 6\,\mu m$, width $= 1\,\mu m$, thickness $= 0.17\,\mu m$,

$$\omega_1 = 1.015 \sqrt{\frac{E}{\rho}} \left(\frac{t}{l^2}\right) = 1.015 \left\{\frac{0.17 \times 10^{-6}}{\left(6 \times 10^{-6}\right)^2}\right\} \sqrt{\frac{E}{\rho}} = 4722.22 \sqrt{\frac{E}{\rho}} \qquad (4.34)$$

For cantilever length $= 10\,\mu m$, width $= 0.6\,\mu m$, thickness $= 0.75\,\mu m$,

$$\omega_1 = 1.015 \sqrt{\frac{E}{\rho}} \left(\frac{t}{l^2}\right) = 1.015 \left\{\frac{0.75 \times 10^{-6}}{\left(10 \times 10^{-6}\right)^2}\right\} \sqrt{\frac{E}{\rho}} = 7500 \sqrt{\frac{E}{\rho}} \qquad (4.35)$$

The angular resonant frequencies are compiled in Table 4.3. For obtaining linear frequencies f, they are divided by 2π.

TABLE 4.3

Resonant Frequencies for Different Cantilever Structures Fabricated Using Various Materials (See Example 4.1)

Material	E/ρ (kg × ms^{-2} m^{-2} kg^{-1} × m^{-3}) = (s^{-2} m^{-2})	$\sqrt{(E/\rho)}$ $\sqrt{(s^{-2}}$ m$^{-2})$ = s^{-1} m^{-1}	w_1 (Radians s^{-1})		
			$l = 200\,\mu m$, $t = 10\,\mu m$	$l = 6\,\mu m$, $t = 0.75\,\mu m$	$l = 10\,\mu m$, $t = 0.75\,\mu m$
Silicon	$1.7 \times 10^{11}/2329 = 7.2993 \times 10^7$	8.54×10^3	2.135×10^6	4.0328×10^7	1.601×10^8
Silicon nitride	$3.1 \times 10^{11}/3290 = 9.422 \times 10^7$	9.71×10^3	2.4275×10^6	4.585×10^7	1.821×10^8
Silicon carbide	$4.1 \times 10^{11}/3100 = 1.323 \times 10^8$	1.15×10^4	2.875×10^6	5.43×10^7	2.156×10^8

4.3.1.2 Deflection Formula

In addition to resonance frequency change, deflection (bending) of the cantilever beam may change due to adsorption-induced surface stress, and it is interesting to enquire the parameters that decide this bending. The governing equation of the beam deflection is as follows:

$$E^* I \left(\frac{d^2 y}{dx^2} \right) = M \tag{4.36}$$

and the boundary conditions are given as

$$y(0) = 0, \quad \frac{dy(0)}{dx} = 0 \tag{4.37}$$

Here the symbol E^* denotes the biaxial modulus, which is the modulus corresponding to biaxial stress, i.e., the stress in a structural member about two perpendicular axes at the same time. E^* takes into consideration the biaxial plane strain conditions associated with thin films and the film–substrate interface. The parameter E^* is defined as

$$E^* = \frac{E}{1 - \mu} \tag{4.38}$$

where
   E is Young's modulus of the beam
   μ is Poisson's ratio (the ratio of transverse contraction strain to longitudinal elongation strain in the direction of stretching force)
   y is the beam deflection
   I is the area moment of inertia

For a rectangular beam,

$$I = \frac{wt^3}{12} \tag{4.39}$$

w and t are the beam width and thickness, respectively, M is the concentrated moment defined as

$$M = \frac{\sigma wt}{2} \tag{4.40}$$

Equation 4.36 is solved by using integration and boundary conditions as follows:

$$y = \frac{Mx^2(1-\mu)}{2EI} \qquad (4.41)$$

For linear analysis of small deflection, the curvature κ of the deflected beam approximately equals d^2y/dx^2. So the curvature κ and the radius of curvature R are given by

$$\kappa = \frac{1}{R} = \frac{M(1-\mu)}{EI} = \frac{6\sigma(1-\mu)}{Et^2} \qquad (4.42)$$

Equation 4.42 is the *Stoney's formula*, serving as a cornerstone for the analysis of curvature-based measurements. Stoney's (1909) formula has been modified by later workers many times for the purpose of evaluating bilayer and multilayer structures with arbitrary layer thickness ratios. Here, it was assumed for the sake of clarity and simplicity that surface stress existed only on the upper surface of the beam. For more generalized cases, σ is substituted by differential surface stress

$$\sigma = \sigma_+ - \sigma_- \qquad (4.43)$$

where σ_+ and σ_- are surface stresses on the upper and lower surfaces, respectively. Also, this substitution does not affect any derivation for linear analysis. Evidently, the modeling of the surface stress effect as a concentrated moment at the free end of the cantilever beam assures that the beam curvature is uniform (constant).

The radius of curvature R is related to the deflection y and length l of the cantilever as

$$\frac{1}{R} = \frac{3y}{2l^2} \qquad (4.44)$$

An equation connecting the cantilever displacement with the differential surface stress is obtained as

$$y = \frac{4l^2\sigma(1-\mu)}{Et^2} \qquad (4.45)$$

Therefore, for remaining parameters remaining constant, the deflection of the cantilever is directly proportional to the adsorption-induced

differential surface stress. Surface stress has units of N m^{-1} or J m^{-2}. An alternative popular form of Equation 4.45 contains the multiplier "3" in place of "4."

Example 4.2

For a silicon cantilever of length = 5 μm, width = 1 μm, thickness = 0.1 μm, $E = 1.5 \times 10^{11}$ N m^{-2}, Poisson's ratio of silicon = 0.17. If the deflection of the cantilever is 5 nm, what is the stress developed in the cantilever beam?

Equation 4.45 gives the stress σ as

$$\sigma = \frac{yEt^2}{\{4l^2(1-\mu)\}} = \frac{5 \times 10^{-9} \times 1.5 \times 10^{11} \times (100 \times 10^{-9})^2}{4 \times (5000 \times 10^{-9})^2 \times (1 - 0.17)} = 0.0904 \text{ N m}^{-1} \quad (4.46)$$

4.3.2 Energy Dissipation and *Q*-Factor of Cantilever

In analogy to other types of resonators, quantification of energy dissipation in MEMS/NEMS is commonly done in terms of the *quality factor* or *Q-factor*. One of the essential features of microcantilevers is that they are mechanical devices (oscillators) that accumulate and store mechanical energy. The *Q*-factor is inversely proportional to the damping coefficient, the ratio of the logarithmic decrement (defined as the natural logarithm of the ratio of two successive maxima of the decaying sinusoid) of any under damped harmonic motion to its period, or total energy lost per cycle of vibration in a microcantilever transducer, and is defined as

$$Q = \frac{2\pi W_0}{\Delta W_0} \quad (4.47)$$

where W_0 and ΔW_0 are, respectively, the mechanical energy accumulated and dissipated in the device per vibration cycle. As the *Q*-factor critically controls both the resonance behavior of any microcantilever and its off-resonance thermal noise, it is an important parameter characterizing MEMS/NEMS sensors operating in resonance as well as static regimes. *Thermal noise* is the noise generated by thermal agitation of electrons in a conductor. The noise power, P, in watts, is given by $P = kT\Delta f$, where k is Boltzmann's constant in joules per Kelvin, T is the conductor temperature in Kelvin, and Δf is the bandwidth in hertz. Thermal noise power, per hertz, is equal throughout the frequency spectrum, it depends only on k and T.

Based on the spectral analysis, an advanced mathematical technique for studying phenomena that occur in cycles, the *Q*-factor is calculated as a ratio of the resonance frequency f_0 to the width of the resonance peak at its half amplitude. Hence, the *Q*-factor is frequently used to represent the degree of the resonance peak sharpness.

What parameters influence the Q-factor of a microcantilever? It depends on a number of intrinsic (belonging to the essential nature or constitution of a thing) and extrinsic (not a basic part or quality) parameters, such as cantilever constructional material, its geometrical shape, and the viscosity (the quantity that describes a fluid's resistance to flow) of the medium in which the cantilever is immersed. Obviously, increased damping of a microcantilever oscillator by the medium translates into lower *Q*-factor values as compared to the same oscillator in vacuum. Models of drag forces, surface friction and pressure, exerted on solid bodies in fluids are used to appraise viscous damping effects.

The other damping mechanisms involve clamping loss and internal friction (the motion-resisting force between the surfaces of particles making up a substance) within the microcantilever. *Clamping loss*, occurring at the joint, grip, or support, has a negligible contribution to the total dissipation in the longer microcantilever possessing high length-to-width and width-to-thickness ratios. However, the ultimate clamping loss minimization is achieved in oscillators with double-paddle or butterfly geometries instead of single-clamped cantilevers or double-clamped bridges. Hence, scientists delving into fundamental studies of intrinsic friction effects in MEMS, frequently perform resonance measurements in double-paddle resonators. *Q*-factors as high as 10^5 were reported for torsional butterfly-shaped resonators fabricated from single-crystal silicon. Torsion refers to deformation caused when one end of an object is twisted in one direction and the other end is held motionless or twisted in the opposite direction.

Internal friction is linked to assorted physical phenomena, especially, thermoelastic dissipation (TED) motion of lattice defects (an energy loss mechanism whereby mechanical energy stored in the cantilever is irrecoverably transferred to the thermal domain; strain gradients lead to temperature gradients in the cantilever, and if these temperature gradients are allowed to relax via heat flow, the mechanical energy is irrecoverable, i.e., TED has occurred), phonon–phonon scattering, and surface effects; phonon is the quantum of acoustic or vibrational energy. In any vibratory structure, the strain field causes a change in the internal energy (the total kinetic and potential energy associated with the motions and relative positions of the molecules of an object, excluding the kinetic or potential energy of the object as a whole) such that its compressed region is heated and extended region becomes cold. Due to the resulting lack of thermal equilibrium between various parts of the vibrating structure, energy is dissipated when irreversible heat flow driven by the temperature gradient occurs; this is TED as pointed out earlier.

The TED limit and phonon–phonon scattering mechanisms correspond to very high *Q*-factors (10^6–10^8), which are scarcely observed experimentally due to the dominant contributions from other dissipation mechanisms afflicting real MEMS devices. The assertion that surface effects constrain *Q*-factors of MEMS oscillators in vacuum is confirmed by annealing the device (treatment

of a device by heating to a predetermined temperature, holding for a certain time, and then cooling to room temperature) and changing its surface in a controlled way. As the thickness of the oscillator decreases, TED becomes a trivial mechanism causing dissipation.

What are the typical magnitudes of Q-factors of cantilevers? Q-factors of MEMS resonators in vacuum are exorbitantly high. However, Q-factors of rectangular microcantilevers in air are typically in the range of 10–1000 while cantilever transducers in aqueous solutions hardly ever have Q-factors above 10.

Very intense viscous damping in liquids makes resonant operation of microcantilevers, and, in turn, measurements of adsorbed mass using microcantilever sensors, somewhat challenging. In order to surmount the difficulties of resonant cantilever operation in liquids, cantilever transducers are used as a part of a self-oscillating system with positive feedback (effect of an action returned to amplify what caused it), e.g., the signal from the microcantilever readout is amplified and fed back to a piezoelectric actuator connected to the microcantilever.

4.3.3 Noise of Cantilever and Its Mass Detection Limit

Since the cantilever resides in a thermal bath with temperature T, the free cantilever is subjected to a random time-dependent Brownian motion around its equilibrium position. This is generally known as *noise. Brownian motion* is the irregular motion in pseudo-random (a process that appears to be random but it is not) or *stochastic* paths (involving chance or probability, not deterministic) of small particles in a fluid that takes place even if the fluid in question is calm. It arises from the thermal motion of the molecules of the fluid.

Noise processes in microcantilever sensors (Albrecht et al. 1991) can be distinguished into: (i) processes intrinsic to the device and (ii) those due to interactions of the cantilever with its environment, e.g., adsorption–desorption noise or noise originating from the readout circuit. Here the focus is on the intrinsic noise mechanisms. *Why?* Because it is these mechanisms that determine the ultimate fundamental limits of the microcantilever sensor performance and thereby establish the lower mass detection boundary of the device.

Relentless efforts have been made to recognize the intrinsic sources of noise in a mechanical system and identify the relationships between parameters of the mechanical system and its noise level. In contrast to a simple harmonic oscillator, a cantilever has different eigenfrequencies (resonance frequencies; eigen = own, particular) ω_n and a mode-dependent spring constant (the restoring force of a spring per unit of length; a measure of how stiff the spring is) D_n. For the nth mode of the cantilever, and for a small frequency interval $\Delta\omega$, on resonance, the average fluctuation $<u_n^2>$ of the sensor is given by (Rast et al. 2000)

$$<u_n^2> = \left(\frac{2k_BT}{\pi}\right)\left(\frac{Q\Delta\omega}{\omega_n D_n}\right) \tag{4.48}$$

where
  k_B is Boltzmann's constant
  T is the absolute temperature
  Q is the quality factor
  D_n is the mode-dependent spring constant

For the first four modes, D_n has the values as under: $D_1 = 3EI/L^3$, $D_2 \approx 121.3D_1$, $D_3 \approx 951.6D_1$, and $D_4 \approx 3654.3D_1$. It is easily surmised that the noise amplitude increases at higher temperatures T and decreases dramatically at higher modes since the mode-dependent spring constant increases. *Prima facie*, it appears that the force sensitivity of an interacting cantilever can be increased by using higher eigenmodes in ambient atmosphere due to the reduction of viscous air damping. But this seems to be only one aspect of the circumstances. The quality factors Q of commercial cantilevers were measured by Rast et al. (2000). The base pressure of the vacuum system was maintained at 10^{-5} mbar throughout the experiment. It was found that for the second harmonic, the Q-value was more than 50% smaller than for the first harmonic. This unequivocally indicated that internal friction of the sensor has a frequency-dependent behavior. Since the Q-factor decreased on the second eigenmode (a normal mode of vibration of an oscillating system), the sensitivity cannot be increased using only higher eigenmodes. To improve the sensitivity, crystalline materials with a small internal friction, such as aluminum or silicon, should be used. For small structures, the surface and bulk represent two independent dissipation channels. A native SiO_2 layer grows on silicon cantilevers when exposed to air. The internal friction of bulk SiO_2 is larger than that of crystalline Si. The sensitivity of the cantilever is therefore appreciably reduced. Promising results are obtained on removing this native oxide layer by HF etching, sputtering, and/or annealing under ultrahigh-vacuum conditions.

   What are the deciding factors for the ultimate mass detection limit of a cantilever? Let us try to answer this question. Upon equilibrating a microcantilever detector with the ambient thermal environment (a thermal bath), a continuous exchange of the mechanical energy accumulated in the device occurs with the thermal energy of the environment. This exchange, dictated by the fluctuation dissipation theorem (a nonequilibrium state may have been reached either as a result of a random fluctuation or an external force, e.g., an electric or magnetic field, and that the evolution toward equilibrium is the same in both cases for a sufficiently small fluctuation), produces spontaneous oscillation of the microcantilever. Hence, the average mechanical energy per mode of cantilever oscillation is defined by thermal energy k_BT. This noise is referred to as thermally induced cantilever noise. Also called

Johnson noise, thermal noise is the random *white* noise generated by thermal disturbance of electrons in a conductor or electronic device. *White noise* is a random signal (or process) with a flat power spectral density meaning that the power spectral density is nearly equal throughout the frequency spectrum. Thus, it is intelligible that any cantilever in equilibrium with its thermal environment has a built-in source of white thermal noise $\psi_{th}(f)$ given by (Albrecht et al. 1991)

$$\psi_{th} = \frac{4m_0 k_B T}{Q} \qquad (4.49)$$

where m_0 is the effective suspended mass of the cantilever. At frequencies much below the resonance frequency, the amplitude of the thermally induced oscillation of a cantilever beam is proportional to the square root of the thermal energy and is expressed as (Lavrik et al. 2004)

$$\sqrt{<\delta z^2>} = \sqrt{\frac{2k_B T B}{\pi k f_0 Q}} \qquad (4.50)$$

where B is the bandwidth requirement. Clearly, lower cantilever stiffness corresponds to higher amplitudes of thermal noise. Consequent upon the dynamic exchange between cantilever mechanical energy and the ambient thermal energy, the actual frequency f of thermally induced cantilever oscillations at any given moment markedly deviates from the resonance frequency f_0. The amplitude of such frequency fluctuations δf_0 due to the exchange between mechanical and thermal energy is (Albrecht et al. 1991)

$$\delta f_0 = \left(\frac{1}{z_{max}}\right)\sqrt{\frac{2\pi f_0 k_B T B}{kQ}} \qquad (4.51)$$

where z_{max} is the amplitude of the cantilever oscillations. Equation 4.51 prophesies increased absolute fluctuation of the resonance frequency δf_0 as the resonance frequency f_0 increases. Nevertheless, relative frequency instability $\delta f_0/f_0$ decreases for higher frequency oscillators (Lavrik et al. 2004)

$$\frac{\delta f_0}{f_0} = \left(\frac{1}{z_{max}}\right)\sqrt{\frac{2\pi k_B T B}{kQ f_0}} \qquad (4.52)$$

By changing the physical dimension of a cantilever, its mass detection limit is affected by many orders of magnitude. For a given cantilever design, the smallest (thermal-noise limited) detectable change in the surface density is

found by combining Equation 4.52 with the equation for fundamental reso-
nance frequency f_0 of the microcantilever:

$$f_0 = \left(\frac{1}{2\pi}\right)\sqrt{\frac{k}{m_0}} \tag{4.53}$$

for a spring constant k and an effective suspended mass m_0 which consists
of both a concentrated and a distributed mass. It is expressed as (Lavrik
et al. 2004)

$$\Delta m_{th} = 8\sqrt{\frac{2\pi^5 k k_B T B}{f_0^5 Q}} \tag{4.54}$$

This equation shows that the mass sensitivity of a cantilever transducer
operating in the resonance mode increases as its dimensions are reduced.
Therefore, cantilever sensors with progressively increased mass sensitivity
are fabricated by simply decreasing the transducer dimensions. As the tech-
nology of nanosize mechanical structures moves forward, nanomechanical
devices approach the gigahertz frequency domain that is already widely dis-
covered with electronic and optical devices.

How are the aforesaid small-dimension cantilevers realized practically? Two
advances have been crucial to breaking the 1 GHz barrier in NEMS (Huang
et al. 2003). First and foremost is the use of silicon carbide epilayers, which
are of comparable density to the usual silicon but are multiple fold stiffer
than silicon and which allow higher frequencies to be attained for structures
of similar geometry. This will be enquired into deeply ahead. Second fac-
tor is the development of balanced, high-frequency displacement transduc-
ers, which enable the ubiquitous passive embedding impedances brought
into play by electrical connections to the macroworld to be nulled. If left
uncontrolled, these parasitic impedances foreshadow the electromechanical
impedance of interest, the signal, in ultrasmall NEMS. The aforementioned
advancements overcome the dual challenge of detecting tiny displacements
(on the scale of femtometers) at microwave frequencies. The characteristic
frequency of NEMS scales upward with decreasing size, but their displace-
ment (when operating linearly) and their electromechanical impedance both
simultaneously scale downward.

Fundamental mechanical resonances for two nominally identical, 3C–SiC
doubly clamped beams, roughly 1.1 µm long, 120 nm wide, and 75 nm thick
were detected at 1.014 and 1.029 GHz by Huang et al. (2003).

The advantages favoring silicon carbide require detailed explanation. SiC
is an excellent material for high-frequency NEMS for the following reasons:
(i) The goal of attaining extremely high fundamental resonance frequencies
in NEMS, while simultaneously preserving small force constants necessary

for high sensitivity, requires pushing against the ultimate resolution limits of lithography and nanofabrication processes. Flexural (a curve, turn, or fold) mechanical resonance frequencies for beams directly depend upon the ratio (E/ρ). The ratio of Young's modulus, E, to mass density, ρ, for SiC is drastically higher than for other commonly used semiconducting materials for electromechanical devices, e.g., Si and GaAs. SiC, given its larger (E/ρ), yields devices that operate at tremendously higher frequencies for a given geometry, than otherwise possible using conventional materials. (ii) For NEMS the Q-factor is governed by surface defects and depends on the device surface-to-volume ratio. SiC possesses excellent chemical stability. This makes surface treatments an option for higher quality factors (Q-factor) of resonance.

4.3.4 Doubly Clamped and Free–Free Beam Resonators

The mathematical analysis carried out for the cantilever is extendable to other MEMS/NEMS structures. For a beam clamped on both sides, the boundary conditions are written as

$$U_y(0,\omega) = 0, \quad \frac{dU_y(0,\omega)}{dx} = 0, \quad U_y(l,\omega) = 0, \quad \frac{dU_y(l,\omega)}{dx} = 0 \qquad (4.55)$$

while for a beam free on both sides, the boundary conditions are stated as

$$\frac{d^2U_y(0,\omega)}{dx^2} = 0, \quad \frac{d^3U_y(0,\omega)}{dx^3} = 0, \quad \frac{d^2U_y(l,\omega)}{dx^2} = 0, \quad \frac{d^3U_y(l,\omega)}{dx^3} = 0 \qquad (4.56)$$

The solution method is similar to the one formulated for the cantilever. The final solution is the same for the clamped–clamped and the free–free beam and only differs from the cantilever in the factor β_i, whose values are also listed in Table 4.1.

Doubly clamped beam resonators based upon NEMS with operating frequencies within the microwave L-band (1–2 GHz) have been achieved. Quality factors of these microwave NEMS resonators decreased as the device frequency was increased. *What could be the causes?* A major possible reason for the decrease of quality factor in these devices is the *clamping loss* intrinsic to the doubly clamped boundary condition.

Wang et al. (2000) explored the use of a free–free boundary condition to reduce this source of acoustic loss for micromachined resonators. Huang et al. (2003, 2005) implemented the in-plane (lateral) free–free beam design strategy to nanoscale mechanical resonators, to explore the possibility of reducing clamping loss. They fabricated pairs of devices consisting of a free–free beam and a doubly clamped beam, very close in designed fundamental-mode resonance frequencies, for comparison in control experiments.

How do the quality factors for the two structures differ? The SiC free–free beam nanomechanical resonators offered extraordinary improvement in quality factor compared to doubly clamped beam design operating at similar frequencies. The free–free beam had a much higher quality factor than the best value from its doubly clamped counterpart. The quality factor of the doubly clamped beam was ~4,500, whereas that of the free–free beam resonator was ~11,000. The dissipation according to clamping loss in such devices increased as the aspect ratio was reduced.

What is the effect of film surface roughness on device performance? Devices were fabricated from films differing widely in surface roughness to investigate the effect of film quality on resonator performance. Devices made from films that had a low-surface roughness (~2 nm or below) operated well in the UHF (300 MHz to 3 GHz)/microwave (3–30 GHz) regime. In contrast, devices made from rougher films (up to ~7.1 nm) were operational only in the VHF (30–300 MHz) range, but not higher. These results firmly indicate that it is critically important to optimize the growth processes to produce ultrasmooth thin films if these films are to be used for nanomechanical resonators having reasonably high-Q values.

By examining devices made from SiC wafers with different roughness, a strong correlation between surface roughness and quality factor was established from experiments.

Example 4.3

Doubly clamped beam was fabricated using silicon carbide. The dimensions were 1000 nm (length) × 125 nm (width) × 75 nm (thickness). Find the resonant frequency.

For a doubly clamped beam, from Equations 4.30 through 4.32 and Table 4.1, the resonance frequency formula is

$$\omega_1 = \left(\frac{22.373}{l^2}\right)\sqrt{\frac{Ewt^3}{12\rho tw}} = 6.46\sqrt{\frac{E}{\rho}}\left(\frac{t}{l^2}\right) \tag{4.57}$$

For silicon carbide, $\sqrt{(E/\rho)} = 1.15 \times 10^4$ s^{-1} m^{-1} (from Table 4.3). Since $t/l^2 = 75 \times 10^{-9}/(1000 \times 10^{-9})^2 = 75,000$ m^{-1}, therefore, $\omega_1 = 6.46 \times 1.15 \times 10^4 \times 75,000 = 5.572 \times 10^9$ rad s^{-1}.

4.4 Electron Tunneling Displacement Nanosensor

Let us now move to displacement nanosensors. The first of these works on the quantum-mechanical phenomenon of electron tunneling. The tunneling transducer (Figure 4.12) is generally realized in the form of a sharp tip placed within a fraction of a nanometer of the moving mechanical element
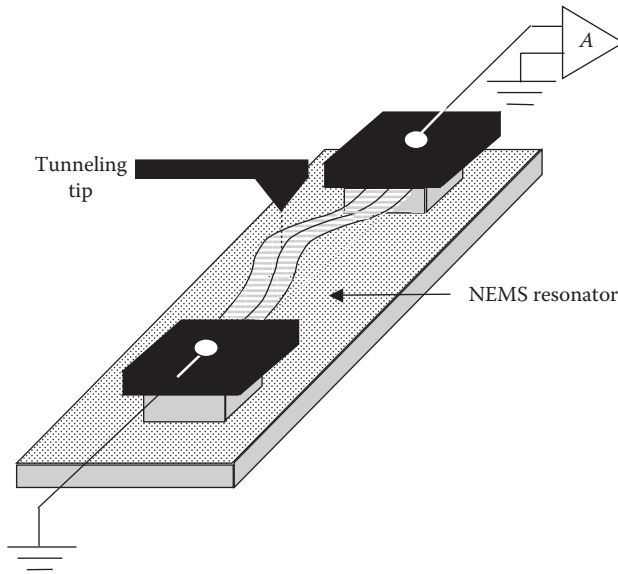
FIGURE 4.12
Electron tunneling displacement nanosensor. (After Ekinci, K. L., *Small*, 1, 786, 2005.)

(Ekinci 2005). By drawing a tunneling current from the surface of the mechanical element, the tip converts the displacement $x(t)$ into an electrical signal. Generally, a tunneling transducer is followed by a transimpedance amplifier (current-to-voltage converter).

The tunnel current i through the junction is related to the DC bias voltage V and the tunnel gap d as

$$i = \rho_s(E_F)V \exp(-2\kappa d) \tag{4.58}$$

There are two important quantities that determine the transduction characteristics: $\rho_S(E_F)$ and κ. The former quantity, $\rho_S(E_F)$, is the local density of electronic states in the test mass. For $V \ll E_F$ where E_F is the Fermi energy (the average energy of electrons in a metal) of the metal in question, $\rho_S(E_F)$, can be assumed to be voltage-independent. The latter, κ, is the decay constant (the decrease of some physical quantity according to the exponential law) for the electron wave function within the gap and

$$\kappa = \frac{\sqrt{2m_e\phi}}{\hbar} \tag{4.59}$$

where
  m_e is the mass of the electron
  ϕ is the approximate work function of the metal
  $\hbar$ is Planck's constant divided by 2π

$\phi \sim 3\text{--}5\,\text{eV}$ and $\phi \gg \text{eV}$ in most cases. Using typical values, one can determine that $\kappa \sim 0.1\,\text{nm}^{-1}$. Keeping in view the exponential dependence of i upon the tunnel gap and the fact that $\kappa \sim 0.1\,\text{nm}^{-1}$, it is irresistibly agreed that a tunneling transducer is extremely sensitive to the displacement of the mechanical element.

4.5 Coulomb Blockade Electrometer–Based Displacement Nanosensor

An electrometer is a high-impedance device for measuring a voltage that draws no current from the source. It is also used to measure a low current (nanoamperes) by passing the current through a high resistance.

4.5.1 Coulomb Blockade Effect

Why does Coulomb blockade take place on a conductor? The Coulomb blockade effect arises because for every additional charge *dq* which is transported to a conductor, work has to be done against the field of already present charges residing on the conductor. Charging an island of capacitance *C* with an electron of charge *e* requires an energy

$$E_c = \frac{e^2}{2C} \tag{4.60}$$

Under these conditions, no additional electron can move into the island unless its energy is raised through an amount E_c by an external voltage

$$V_c = \frac{E_c}{e} = \frac{e}{2C} \tag{4.61}$$

What is Coulomb blockade regime? For voltages $V < V_c$, at zero temperature, no current can flow into the island and the system is said to be in *Coulomb blockade regime*.

   How is the Coulomb blockade effect modulated? The effect is modulated by coupling a small capacitor into the island forming a three-terminal device called the *Coulomb blockade electrometer* (White 1993). Thus the Coulomb blockade electrometer consists of two tunnel junctions (extremely thin potential barriers to electron flow, so that the transport characteristic, the current–voltage curve, is primarily governed by the quantum-mechanical tunneling process which permits electrons to penetrate the barrier) of capacitances C_1 and C_2,
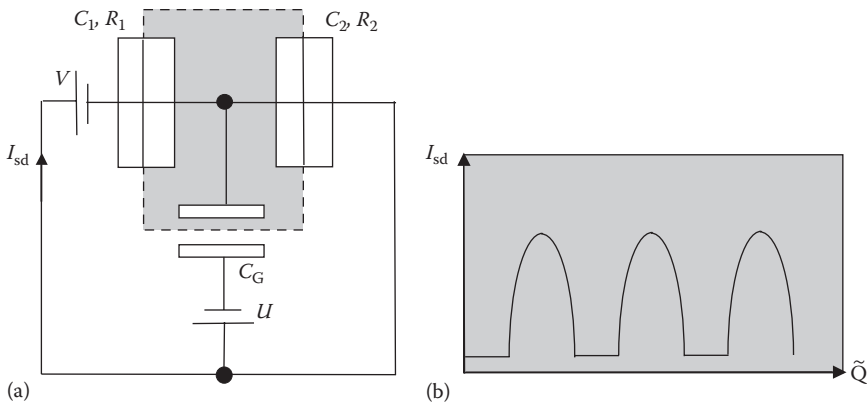
FIGURE 4.13
(a) Coulomb blockade electrometer and (b) its response to varying charge. (After White, J. D., *Jpn. J. Appl. Phys.*, 32, L1571, 1993.)

as shown in Figure 4.13, where the Coulomb island is marked by dashed line. The electrometer is completed by a capacitor C_G weakly coupling into the central region of the device, the island, biased by a voltage U with the island externally biased by a supply V. The capacitor C_G constitutes the gate of the device.

A background charge q_0 on the island produces generally a nonintegral charge offset. *How is this background charge induced?* The background charge is induced by stray capacitances that are not shown in the circuit diagram and impurities located near the island, which are practically always present. The effect of the gate electrode is that the background charge q_0 can be changed at will, because the gate additionally polarizes the island.

Consider the situation when both the gate and island voltages are zero. Then electrons do not have enough energy to enter the island and current does not flow. For an electron to hop on to the island, its energy must equal the Coulomb energy $e^2/2C$. As the bias voltage between the source and drain is increased, the energy in the system reaches the Coulomb energy and an electron can pass through the island. The critical voltage needed to transfer an electron onto the island, equal to e/C, is called the *Coulomb gap voltage*.

Now imagine that the bias voltage is kept below the Coulomb gap voltage. When $U=0$ and $V<E_c/e$, transport through the island is impeded by Coulomb blockade because there is an energy gap between the states where the island has a charge $-Ne$ and $-(N+1)e$. By changing U, an extra charge $\bar{Q} = C_G U$ is induced on the island, which is added to $-Ne$. It is possible to make these states degenerate thereby lifting the Coulomb blockade. Thus, the charging energy may be overcome by changing the source–drain voltage as well as by changing the gate voltage.

On applying suitable gate bias, the device starts conducting with a maximum source–drain current $I_{\text{sd (maximum)}}$ given by

$$I_{\text{sd(maximum)}} \sim \frac{V}{2R} \tag{4.62}$$

A closer introspection reveals that the current is maximum at a charge

$$\tilde{Q} = \frac{n|e|}{2} \tag{4.63}$$

and when $\tilde{Q} = |e|$, the device returns to the state equivalent to zero gate bias. Again, there is an energy gap between the states $-(N+1)e$ and $-(N+2)e$. Thus a cyclic process takes place with the current I_{sd} oscillating between low and high values. The period of oscillation in $\tilde{Q}$ is $|e|$. This implies that the conductivity of the island can be significantly varied by changing $\tilde{Q}$, thereby enabling measurements of $\tilde{Q}$ changes at a level lower by several orders of magnitude than the electronic charge e.

Thus far, $\tilde{Q}$ was changed by altering the voltage U applied to the gate. There is another technique by which $\tilde{Q}$ can be changed. *How?* By changing the value of C_G at constant gate bias U by varying the separation between the two plates of C_G, e.g., by keeping the plate near the island fixed and transferring the other plate relative to it. This produces an identical effect on $\tilde{Q}$.

How does the electrometer sense displacement? The electrometer acts as a motion transducer because it is able to sense the movement of the capacitor plate. The smallest displacement that can be sensed by the electrometer ($\delta X_{\text{minimum}}$) is expressed as

$$\delta X_{\text{min}} = \frac{X \delta Q_{\text{minimum}}}{C_G U} \tag{4.64}$$

where gate is formed by suspending a conducting plate above the island at a height X and $\delta Q_{\text{minimum}}$ is the minimum charge that can be detected by the electrometer.

When $C_g \ll$ minimum (C_1, C_2), the charge $\delta Q_{\text{minimum}}$ is estimated as (White 1993)

$$\delta Q_{\text{minimum}} = 5.4e \left\{ \frac{k_B T}{(e^2 / C_{\text{minimum}})} \right\}^{0.5} \sqrt{R C_{\text{minimum}} \Delta f} \tag{4.65}$$

where

   C_{minimum} is the capacitance of the smallest tunnel junction which contemporary manufacturing technology can make
   Δf is the relevant bandwidth

Combination of Equations 4.64 and 4.65 yields the displacement sensitivity

$$\frac{\delta X_{\text{minimum}}}{\sqrt{\Delta f}} \approx \left(\frac{21.6}{e}\right)\sqrt{k_{\text{B}}TR}XC_{\text{minimum}} \tag{4.66}$$

assuming that the biasing point for the device is $\tilde{Q} = e/4$. Using electron beam technology, $C_{\text{minimum}} \sim 10^{-16}$ F. Taking X to be $10\,\mu\text{m}$, the displacement sensitivity is $5 \times 10^{-12}\,\text{mHz}^{-1/2}$ at $T = 10\,\text{mK}$.

4.5.2 Comparison with Tunneling Sensors

Following are the points of dissimilarity between the two:

1. Although electron tunneling sensors have a high sensitivity at least by two orders of magnitude, they have a dynamic range of the extent of tip-surface distance around 5 Å. Larger dynamic range requires integration of the tip with a translational component so that the final resolution is determined by the tip-translator combination. In contrast, the dynamic range of an electrometer-based sensor is the gate-island distance, which is much larger than the bare tunneling tip.

2. Moreover, establishing a reliable tunnel junction contact is not easy outside an ultra-high-vacuum environment. Coulomb blockade sensor does not suffer from this limitation.

3. The surface from which the current tunnels off is never flat on the atomic scale. This means that a large change in tunneling current may occur when the tunneling tip sweeps across a small vertical step downward or upward, showing that this kind of sensor is not a one-dimensional motion sensor. The same is not true for Coulomb blockade device.

4.6 Nanometer-Scale Displacement Sensing by Single-Electron Transistor

The SET is the most sensitive electrometer, with a demonstrated sensitivity below $10^{-5}\,\text{eHz}^{-1/2}$. Presently, a promising approach for displacement sensing is to use a SET capacitively coupled to a flexural beam (a beam subjected to bending action) resonator (Figure 4.14).

Position measurements of an oscillator are ultimately limited by quantum mechanics, where "zero-point motion" fluctuations in the quantum ground
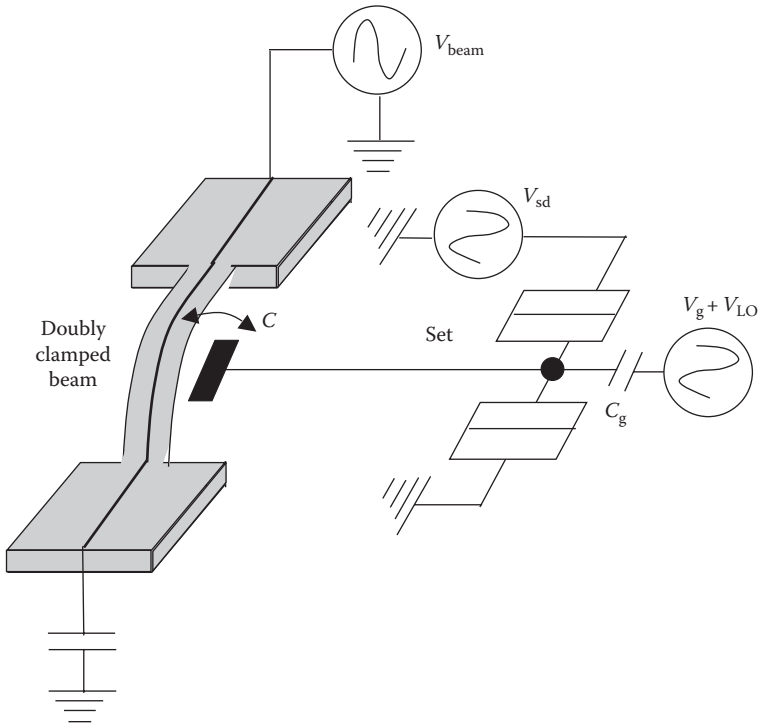
FIGURE 4.14
Capacitive coupling of the electrode on resonator flexural beam to SET gate electrode. (After Knobel, R. G. and Cleland, A. N., *Nature*, 424, 291, 2003.)

state combine with the uncertainty relation to give way to a lower limit on the measured average displacement. Average total energy of a classical simple harmonic oscillator (any physical system that is bound to a position of stable equilibrium by a restoring force or torque proportional to the linear or angular displacement from this position), in equilibrium with its environment at temperature T, is $k_B T$. The position of the oscillator fluctuates continuously with a root-mean-square displacement amplitude given by (Knobel and Cleland 2003)

$$\delta x = \sqrt{\frac{k_B T}{m \omega_0^2}} \tag{4.67}$$

for an oscillator of mass m and resonant frequency $f_0 = (\omega_0/2\pi)$. The classical displacement amplitude can be made arbitrarily small by reducing the temperature. Quantum mechanics has two implications here: (i) The quantized nature of the oscillator energy yields an intrinsic fluctuation amplitude, the "zero-point motion"

$$\delta x_{zp} = \sqrt{\frac{\hbar}{2m\omega_0}} \tag{4.68}$$

which is achieved for temperatures T well below the energy quantum $T \ll T_Q \equiv \hbar\omega_0/k_B$. (ii) The instrument used to measure the position of the oscillator will unavoidably perturb it, further limiting the possible measurement resolution, as quantified by the *Heisenberg Uncertainty Principle* (increasing the accuracy of measurement of one observable quantity increases the uncertainty with which another conjugate quantity may be known).

An implementation that allows near quantum-limited sensitivity uses a SET as a displacement sensor. The exquisite charge sensitivity of the SET at cryogenic temperatures is exploited to measure motion. Knobel and Cleland (2003) demonstrated an ultrasensitive, potentially quantum-limited displacement sensor based on a SET, enabling to read out the motion of a nanomechanical resonant beam at its resonant frequency. For a model cantilever and SET, applying the optimal bias voltage yields a displacement sensitivity of 4×10^{-16} mHz$^{-1/2}$, approaching the sensitivity needed to measure quantum effects.

What is the construction of SET and how does it work? The SET consists of a conducting island separated from leads by low-capacitance, high-resistance tunnel junctions. The current through the SET is modulated by the charge induced on its gate electrode, with a period e, the charge of one electron.

How is SET applied for detecting the motion of a nanomechanical resonator? By capacitively coupling the gate of the SET to metal electrode on a flexural beam of the resonator and biasing the electrode at a constant voltage V_{beam}. In this scheme, the capacitance C between the SET and the beam has a coupled charge $q = V_{beam} C$. As the beam vibrates in the x direction, in the plane of the device, the resulting variation in capacitance modulates the charge induced on the SET, $\Delta q = V_{beam} \Delta C$, changing the SET source–drain current. When the voltage V_{beam} is increased, the charge modulation Δq and the sensitivity to the resonator motion increases. However, the source–drain current is due to the stochastic flow of electrons through the SET, so the voltage of the center island voltage fluctuates randomly. This causes a fluctuating "back-action" force on the beam. Increasing the voltage applied to the beam leads to larger coupled charge signals, but also increases the back-action coupling between the SET and the beam. The force increases as V_{beam} increases, resulting in a voltage for which the total noise is minimized. The device had a displacement sensitivity of 2.0×10^{-15} mHz$^{-1/2}$ at the 116.7 MHz resonant frequency of the mechanical beam at a temperature of 30 mK, which was limited by the noise in the conventional electronics. The sensitivity was roughly two orders of magnitude larger than the quantum limit for this oscillator.

4.7 Magnetomotive Displacement Nanosensor

Cleland and Roukes (1996) introduced the magnetomotive (designating or of a force that gives rise to magnetic flux) displacement detection technique into the NEMS domain. This scheme is based upon movement of a conducting nanomechanical element through a uniform static magnetic field (Figure 4.15). The resultant time-varying flux generates an induced electromotive force (EMF) in the loop, which, in turn, is picked up by the detection circuit (Ekinci 2005).

4.8 Piezoresistive and Piezoelectric Displacement Nanosensors

These sensors do not detect the actual displacement. *How do piezoresistive/ piezoelectric sensors sense displacement*? Piezoresistive and piezoelectric detection are both sensitive to the strains generated inside a material during its motion. A *piezoresistive material* is one that exhibits a change in resistivity due to strain. Likewise, a *piezoelectric material* becomes electrically polarized in the event of a deformation.
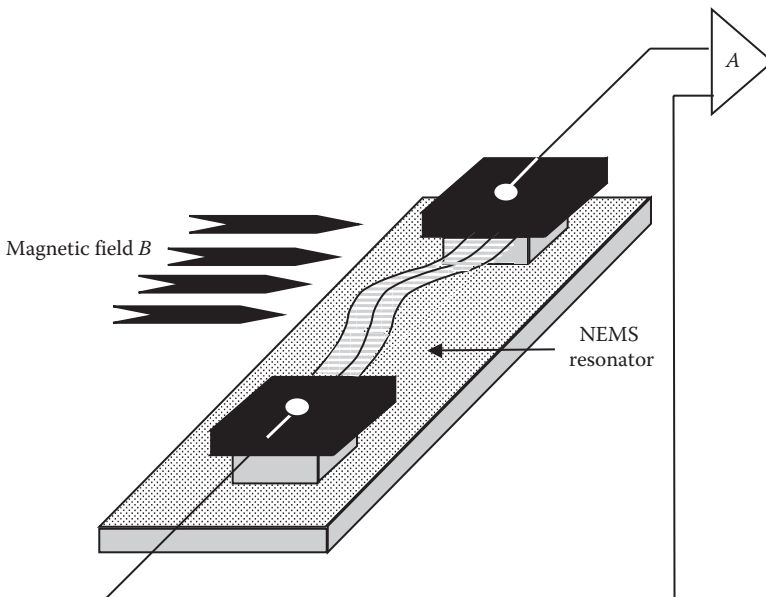
FIGURE 4.15
Magnetomotive displacement nanosensor. (After Ekinci, K. L., *Small*, 1, 786, 2005.)

Piezoresistive sensing is realized by detecting the resistance changes through a piezoresistive NEMS device upon actuation (Ekinci 2005). There are, however, some obstacles in this method, e.g., in a P-type Si doubly clamped beam, an optimistic estimate for the resistance change ΔR, as a function of the displacement, the beam length l, and the resistance R, is $\Delta R/R \sim x^2/l^2$, indicating that $\Delta R_{max}/R \sim 1/100$. In view of the already high resistance of a doped semiconductor beam approaching $R \sim 10 \, k\Omega$, such a small resistance change is likely to be obscured at high resonance frequencies by the inescapably accompanying parasitics.

Polarization fields created by the strain fields within the piezoelectric nanomechanical element, as it undergoes motion, form the basis of the piezoelectric scheme. The detection is realized by precisely measuring the potential drop across the strained device during its motion. In most materials, the piezoelectric coefficients linking the strain to the polarization field are $\sim 10^{-12}$–$10^{-9} \, mV^{-1}$, suggesting that the piezoelectric voltage generated across a nanoscale device is very small. The high source impedance of the generated (piezoelectric) signal makes detection difficult at high frequencies.

Knobel and Cleland (2002) proposed piezoelectric detection in NEMS using a SET (Figure 4.16). Here, the gate of the SET was placed at a position of maximum strain, and detection was realized by monitoring the consequential conductance changes in the current-biased SET. In other words, the resonator was fabricated from a piezoelectric material, such as GaAs,
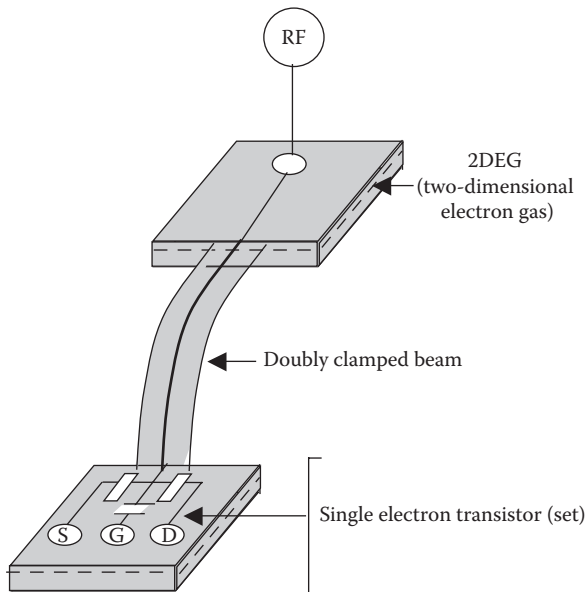
FIGURE 4.16
Piezoelectric displacement sensing using SET. (After Knobel, R. and Cleland, A. N., *Appl. Phys. Lett.*, 81, 2258, 2002.)

AlGaAs, or AlN, and the SET was configured to sense the piezoelectric voltage developed when the beam flexes. A noticeably higher displacement sensitivity was achieved here than by using capacitive displacement sensing, primarily due to the strong piezoelectric coupling strength (Figure 4.8). Knobel and Cleland (2002) calculated a noise figure of 5×10^{-17} mHz$^{-1/2}$, for a 1 GHz GaAs resonator, dominated equally at peak sensitivity by the current and back action noise of the SET. Electromechanical coupling describes the conversion of energy by the element from electrical to mechanical form or vice versa. Coupling coefficient is a dimensionless number related to the effectiveness of electrical to mechanical energy conversion in piezoelectric devices.

Piezoelectric detection has another attractive feature because piezoelectric signals scale favorably to the small, stiff resonators needed to approach the regime where $\hbar\omega_1 \geq k_B T$. In both the capacitive and piezoelectric schemes, the detection of deviations from the classical motion is difficult for nanomechanical resonators with $\omega_1/2\pi < $ GHz.

4.9 Optical Displacement Nanosensor

Optical interferometry techniques (Krishnan et al. 2007), notably path-stabilized Michelson interferometry (the Michelson interferometer produces interference fringes by splitting a beam of monochromatic light into two paths so that one beam strikes a fixed mirror and the other a movable mirror. When the reflected beams are recombined, an interference pattern results) and Fabry–Perot interferometry (this interferometer utilizes multiple reflections between two closely spaced partially silvered surfaces; part of the light is transmitted each time reaching the second surface, and resulting in multiple offset beams which interfere with each other, producing an interference pattern with extremely high resolution, somewhat like the multiple slits of a diffraction grating increase its resolution), have been extended into the NEMS domain. In Fabry–Perot interferometry (Figure 4.17), the optical cavity formed within the sacrificial gap of the NEMS, between the NEMS surface and the substrate, varies the optical signal and directs it onto a photodiode (a semiconductor diode in which the reverse current varies with illumination) with the movement of the NEMS device in the out-of-plane direction.

In path-stabilized Michelson interferometry, a tightly focused probing laser beam of wavelength λ is focused at the center of a doubly clamped beam through an objective lens (the lens of a microscope at the bottom near the sample). It is reflected from the surface of the moving NEMS device and undergoes interference with a stable reference beam.

In both the aforementioned techniques, strong diffraction effects dominate as the relevant NEMS dimensions are shrunk below the optical wavelength
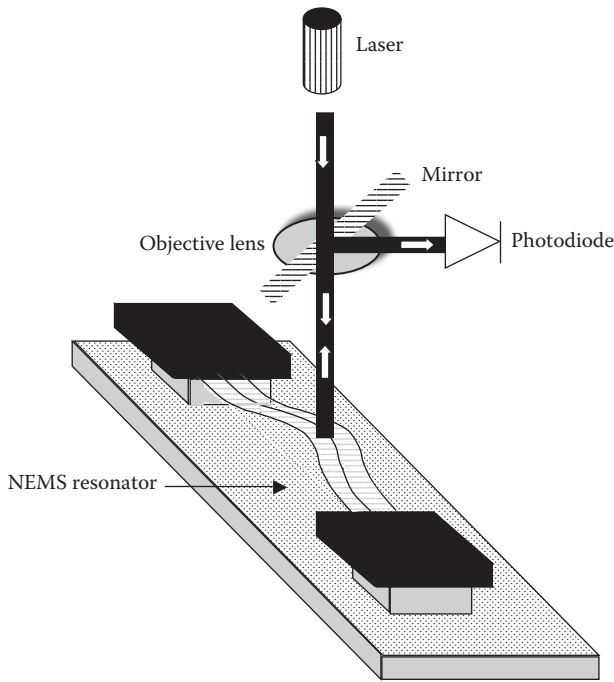
FIGURE 4.17
Optical displacement nanosensor. (After Ekinci, K. L., *Small*, 1, 786, 2005.)

used. The light gathered by the probing lens is weakened to a small fraction of the incoming light by strong scattering.

4.10 Femtonewton Force Sensors Using Doubly Clamped Suspended Carbon Nanotube Resonators

Thus far resonators constructed from silicon, silicon carbide, and related materials were discussed. *What about building resonators using CNTs?* The CNT resonator (Li and Chou 2003, Sazonova et al. 2004, Anantram and L'eonard 2006, Li and Chou 2006) shown in Figure 4.18a consists of a CVD-grown nanotube suspended over a trench etched in silicon dioxide between two metal electrodes, with small portions of CNT overlying the oxide on both sides of the trench. The nanotube adheres to the oxide by clamping at the suspension points.

What does the configuration look like? It is essentially a three-terminal transistor-like arrangement in which the nanotube is the channel between its two
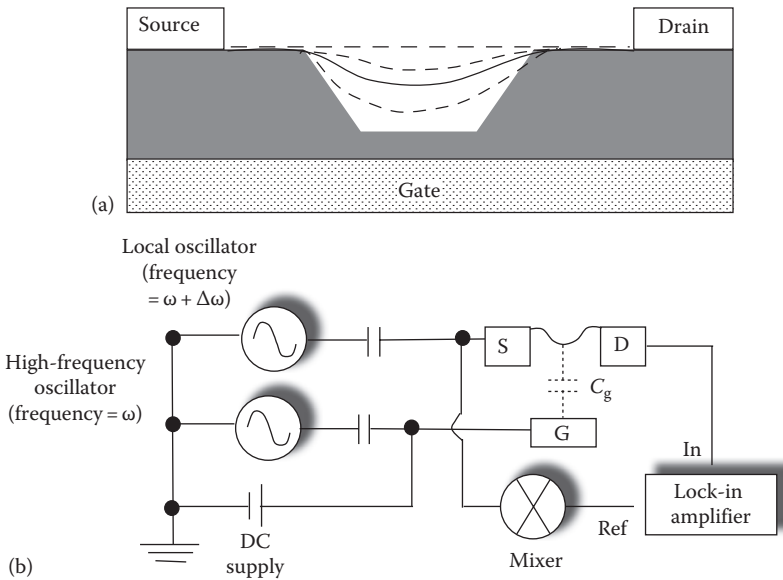
FIGURE 4.18

(a) A CNT resonator. (b) Layout of experimental setup for measurement of resonance frequency of CNT resonator. (After Sazonova, V. et al., *Nature*, 431, 284, 2004.)

metal contacts. The current in the channel is modulated by the voltage applied to the gate through capacitive coupling.

The measurement is done in a vacuum chamber at pressures below 10^{-4} Torr by actuating the nanotube motion through electrostatic interaction with the gate electrode. The applied gate voltage V_g induces an additional charge $q = C_g V_g$ on the nanotube, where C_g is the capacitance between the gate and the tube. The mutual attraction between the charge q and its opposite charge $-q$ on the gate produces a downward electrostatic force on the nanotube bending the tube towards the gate.

What are the components of the gate voltage and what roles do they play? The gate voltage has two components: a static (DC) component and a small time-varying (AC) component. The DC voltage $V_{g(DC)}$ produces a static force on the nanotube that is used to vary its tension. The alternating voltage $V_{g(AC)}$ produces a periodic electric force but the force is *always attractive in nature*.

How does the nanotube vibrate? When $V_{g(AC)} = -V_{g(DC)}$, i.e., the two voltages are equal in magnitude but opposite in sign, the nanotube returns to its previous upward position by virtue of its elastic resilient property because the downward electrical force ceases to act during this period. But when $V_{g(AC)}$ changes its direction or magnitude, a net downward force again starts acting on the nanotube. Thus the nanotube is either pulled downward or there is no force acting on it. During the later period, the nanotube recovers its original

upward position. In this way, the nanotube moves up and down swinging into continuous oscillations.

When the driving frequency matches the natural frequency of the tube, the displacement of the nanotube becomes large due to resonance. *How is the resonance frequency determined?* To find the resonance frequency, the resonance condition must be detected. It must be noted that the change in induced charge on the nanotube (δq) = change in the charge caused by gate voltage variation ($\delta q_1 = C_g \delta V_g$) + change in the charge due to alteration in capacitance between the nanotube and gate arising from the continuous distance variation between them during vibration of the nanotube ($\delta q_2 = \delta C_g V_g$). The resulting AC current flowing in the nanotube is expressed as

$$i_{\text{nanotube}} = \frac{\delta q_1}{\delta t} + \frac{\delta q_2}{\delta t} = C_g \left(\frac{\delta V_g}{\delta t} \right) + V_g \left(\frac{\delta C_g}{\delta t} \right) \tag{4.69}$$

Only the second term is related to the distance between the nanotube and gate, i.e., the amplitude of oscillations. The first term is independent of this amplitude. When ($\delta C_g / \delta t$) is maximum corresponding to maximum displacement of nanotube from its equilibrium position, the current registers peaks in both directions. Thus the change in nanotube current with frequency of AC field is observed and the frequency at which this current swing is maximum pertaining to the largest amplitude is the resonant frequency. Taken as a whole, the resonator exhibits source–drain AC conductance enhancement when the driving frequency of the gate matches the resonant frequency of mechanical oscillations of the nanotube.

How is the resonance condition ascertained practically? To detect the current or conductance change of the nanotube, it is used as a mixer (Figure 4.18b). This method helps in avoiding complications due to capacitive currents between the gate and the drain electrodes. The current equals the product of the AC voltage on the source electrode and the modulated nanotube conductance. The measured current response is fitted to a Lorentzian function (singly peaked function used for fitting of raw data) with a normalized linewidth, a resonant frequency = 55 MHz, and an appropriate phase difference between the actuation voltage and the force on the nanotube.

The resonant frequency increases as the DC gate voltage rises. The DC gate voltage is adjusted to tune the resonant frequency. Nanotubes of dissimilar diameters and lengths also yield different resonant frequencies because the oscillator masses differ.

An important parameter characterizing the oscillator is the quality factor Q, the ratio of the energy stored in the oscillator to the energy lost per cycle owing to damping. It is in the range of 40–200 with no observed frequency dependence. Maximizing Q is important for most applications.

To use a nanotube as a force/strain sensor, it is noted that the resonant frequency of a resonator is closely related with its dimensions. When a carbon
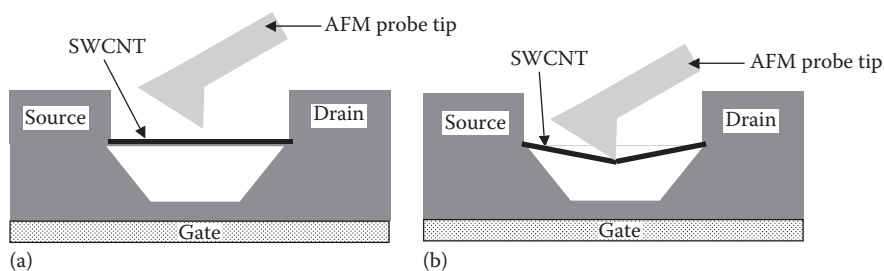
FIGURE 4.19
Using an AFM tip to apply a force on SWCNT: (a) no force applied on the tube and (b) a force applied on the tube. (After Minot, E. D. et al., *Phys. Rev. Lett.*, 90, 156401, 2003.)

nanotube is deformed, the tube length and tube diameter are changed, and its resonant frequency changes accordingly. There exists a relationship between the shift of resonant frequency of a nanotube resonator and the force the nanotube is subjected to. This is the underlying principle for carbon nanotube–based force sensors. Thus, the key issue of study is to establish the relationship between the resonant frequency shift and the applied force.

How is force applied on the nanotube? An AFM tip can be used to simultaneously vary the force on CNT and to electrostatically gate the tube (Figure 4.19) (Minot et al. 2003). The ensuing section will illustrate two configurations of CNT force sensor in which a cantilever and a bridge structure are used to apply the force.

What is the ultimate force sensitivity of the sensor? The ultimate limit on force sensitivity is prescribed by the thermal vibrations of the nanotube. This force sensitivity is 20 aNHz$^{-1/2}$ for typical parameters; atto $= 10^{-18}$. The observed sensitivity $= 1$ fNHz$^{-1/2}$ was 50 times lower than this limit; femto $= 10^{-15}$. *Why was this sensitivity lower?* This is probably due to the low values of transconductance (the ratio of change in current to change in voltage that caused it) for the measured nanotubes at room temperature. At low temperatures (~ 1 K), the sensitivity increases by orders of magnitude due to the higher transconductance of the nanotube.

4.11 Suspended CNT Electromechanical Sensors for Displacement and Force

In a series of papers, researchers at ETH Zurich reported electromechanical sensors based on CNTs (Stampfer et al. 2005, Stampfer et al. 2006b,c,d). A contacted, suspended single-walled carbon nanotube (SWCNT) is fixed underneath a freestanding cantilever (Figures 4.20a and 4.21a) or bridge (Figures 4.20b and 4.21b). On applying an external out-of-plane force on the

(a) (b)
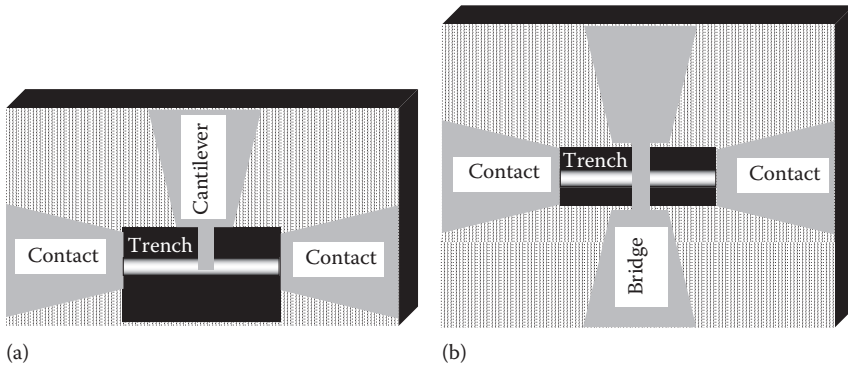
FIGURE 4.20
Configurations for building force sensors by applying force over: (a) cantilever and (b) bridge. (After Stampfer, C. et al., *The 13th International Conference on Solid-State Sensors, Actuators and Microsystems*, Seoul, Korea, June 5–9, 2103, 2005; Stampfer, C. et al., *IEEE Sensors J.*, 6, 613, 2006.)

(a)

(b)
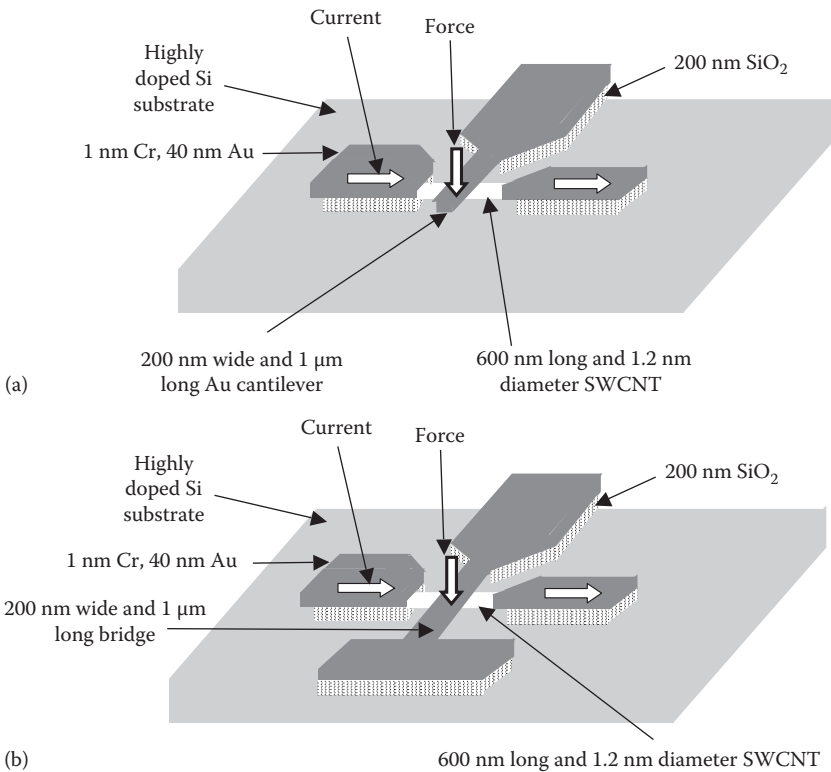
FIGURE 4.21
Suspended CNT-based electromechanical force sensors with: (a) cantilever and (b) bridge. The gap between the CNT and the underlying substrate is provided by oxide layer thickness. (After Stampfer, C. et al., *The 13th International Conference on Solid-State Sensors, Actuators and Microsystems*, Seoul, Korea, June 5–9, 2103, 2005; Stampfer, C. et al., *IEEE Sensors J.*, 6, 613, 2006.)

cantilever or bridge, it undergoes deflection producing a mechanical deformation of the underlying SWCNT. The structural change of the SWCNT yields a change of conductance, which is electrically measured. The conductance changed by a factor of 3 for approximately 120 nN of force for a metallic SWCNT with 5 MΩ base resistance.

An AFM tip introduces the external out-of-plane force for device actuation. In practical application, the cantilever can act as a support spring connected to a seismic mass (the mass that converts the acceleration to spring displacement is referred to as the test mass or seismic mass). Deflection of the seismic mass, e.g., by an acceleration, causes a mechanical deformation of the SWCNT leading to an electrical signal.

Minot et al. (2003) employed an AFM tip to simultaneously vary the CNT strain and to electrostatically gate the tube. They found that, under strain, the conductance of the CNT can increase or decrease, depending on the tube. By using the tip as a gate, they showed that this is related to the increase or decrease in the bandgap of a CNT under strain.

V_{tip} was swept ∼3 times a second over a range of a few volts as strain σ was slowly increased. A CNT is either metallic or semiconducting depending on the orientation between the atomic lattice and the tube axis. As the metallic tube was strained, an asymmetric dip centered at $V_{tip} \sim 1\,V$ developed in $G - V_{tip}$ (*G* is the conductance of the nanotube). On the other hand, the semiconducting tube showed an increase of *G* with strain and a reduction in the asymmetry of the dip. The maximum resistance R_{max} (σ) for each sweep of V_{tip} as a function of strain is fitted to the functional form:

$$R_{max}(\sigma) = R_0 + R_1 \exp(\beta\sigma) \qquad (4.70)$$

where β, R_0 and R_1 are fitting parameters.

β, the exponential fitting parameter, is related to the strain dependence of the gap:

$$\frac{dE_{gap}}{d\sigma} = \beta kT \qquad (4.71)$$

They showed that strain can open a bandgap in a metallic CNT and modify the bandgap in a semiconducting CNT. Theoretical work predicts that bandgap changes can range between ∼100 meV per 1% stretch, depending on CNT chirality (Yang et al. 1999). They obtained $dE_{gap}/d\sigma = -53$ meV for the semiconducting tube and $dE_{gap}/d\sigma = +35$ meV for the metallic tube. They also estimated chiral angles φ ∼ 19° and 23° for the two tubes, respectively. These results can be understood from the relation for the rate of change of bandgap with strain, $dE_{gap}/d\sigma$:

$$\frac{dE_{gap}}{d\sigma} = \text{sign}(2p+1)3t_0(1+\nu)\cos 3\phi \qquad (4.72)$$

where
$t_0 \sim 2.7$ eV is the tight-binding overlap integral
$\nu \sim 0.2$ is the Poisson ratio
ϕ is the CNT chiral angle
$p = -1, 0,$ or 1 such that the wrapping indices, n_1 and n_2, satisfy

$$n_1 - n_2 = 3q + p \qquad (4.73)$$

where q is an integer. The tight binding model or TB model is an approach to the calculation of electronic band structure employing an approximate set of wave functions based upon superposition of wave functions for isolated atoms situated at each atomic site. The maximum value of $|dE_{gap}/d\sigma|$ is $3t_0$ $(1 + \nu) = 100$ meV, similar in magnitude to $dE_{gap}/d\sigma$ of typical bulk semiconductors. Note that half of all semiconducting CNTs $(p = 1)$ will have $dE_{gap}/d\sigma > 0$, while the other half $(p = -1)$ have $dE_{gap}/d\sigma < 0$.

To model the resistance $R_{max}(\sigma)$ associated with this conductance minimum, they considered transport by thermal activation, neglecting tunneling across the depleted region. Electrons with energy E such that $|E - E_F| > E_{gap}$ cross the barrier with a transmission probability $|t|^2$, while those with $|E - E_F| < E_{gap}$ are reflected; transmission probability is the probability that a particle will pass through a potential barrier, that is, through a finite region in which the potential energy of the particle is greater than its total energy.

The low-bias resistance of the device is then

$$R_{total} = R_S + \left(\frac{1}{|t^2|}\right)\left(\frac{h}{8e^2}\right)\left\{1 + \exp\left(\frac{E_{gap}}{kT}\right)\right\} \qquad (4.74)$$

where
the first term R_S is the resistance in series with the junction due to the metal–CNT contacts, etc.
the second term is the resistance of the junction region

From Equation 4.72, we have

$$E_{gap} = E_{gap}^0 + \left(\frac{dE_{gap}}{d\sigma}\right)d\sigma \qquad (4.75)$$

Additional knowledge about the device can be gained from the fitting parameter $R_1 \sim 49h/e^2$, and, hence, transmission probability $|t|^2 = 0.25$. Using an estimate of $|t|^2 = 0.25$ from the preceding text, they inferred $E_{gap}^0 = 160$ meV. This inferred energy gap corresponded to a tube with diameter $d = 4.7$ nm.

Displacement sensing is also obtained from conductance changes. If the nanomechanical structure, of which displacement is to be measured, undergoes a mechanical deflection, the SWCNT is mechanically deformed (mainly axially stretched) which leads to a significant change in the conductance of the SWCNT. An effective differential sensor sensitivity of $1016\,k\Omega$ at a deflection $= 35\,nm$ was obtained for the metallic SWNT giving a relative differential resistance sensitivity $= 27.5\%$ per nm. The piezoresistive gauge factor was 2900.

4.12 Membrane-Based CNT Electromechanical Pressure Sensor

When pressure is applied to a resistor, strain is induced and the resistance of the resistor changes (the piezoresistive effect). The characteristic of a piezoresistor is modeled by its gauge factor (G) defined as

$$G\varepsilon = \frac{\Delta R}{R} \qquad (4.76)$$

where
 ε is the applied strain
 ΔR is the resistance change due to applied strain
 R is the original resistance

Gauge factor (G or GF) also called strain factor of a strain gauge is the ratio of relative change in electrical resistance to the mechanical strain ε. When CNT sensor is used as piezoresistor, the bias current should be limited to a small value so that self-heating is minimized. The threshold value of this bias current varies from sensor to sensor.

The pressure sensor (Figure 4.22) consists of an ultrathin atomic layer deposited (ALD) circular alumina (Al_2O_3) membrane to the surface of which an SWCNT is adhered by van der Waal's forces and clamped in place by two metal electrodes (Li and Chou 2004, Stampfer et al. 2006a). The pressure sensor design is based on circular membranes in contrast to rectangular membranes mainly because of the isotropy (identical in all directions) of the strain in the central area of the membrane. Assuming that the SWCNT experiences the same stretching as the membrane, the device works as a strain gauge similar to the doped-silicon strain gauges used in MEMS pressure sensors.

The pressure sensor has been tested for a pressure range of 0–130 kPa (standard atmospheric pressure is $101{,}325\,Pa = 101.325\,kPa = 1{,}013.25$ mbar $= 760$ Torr). For small pressures (up to $\Delta p \sim 70\,kPa$), a monotonic increase in
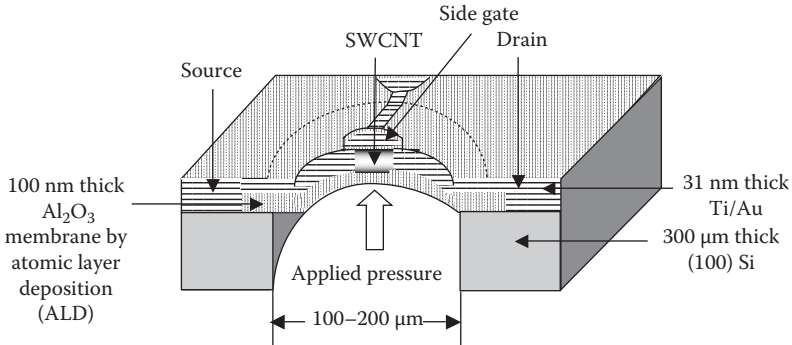
FIGURE 4.22
CNT-on-Al_2O_3-membrane type pressure sensor. (After Stampfer, C. et al., *Nano Lett.* 6, 233, 2006a.)

resistance (consistently increasing and never decreasing in value) with pressure was observed. For larger pressures, there was a non-monotonic anomaly. However, the piezoresistive gauge factor of the metallic SWCNT = 200 ± 8, slightly exceeded the value of doped-Si strain gauge (200).

4.13 Tunnel Effect Accelerometer

4.13.1 Principle of Motion Detection

As already explained in Section 1.19, the STM is a displacement nanosensor utilizing the extreme sensitivity of tunneling current to variations in electrode spacing. Several methods have been developed for implementation of position sensors based on electron tunneling. These methods help in position detection with sub-Angstrom resolution using simple mechanical structure and control circuits. Position or displacement detection, in turn, constitutes the basis of acceleration, force, and pressure measurements, e.g., the acceleration of a body is determined from the displacement of a "proof mass" when the sensor structure undergoes acceleration.

What are the available methods for displacement measurement besides electron tunneling? What are their shortcomings? The methods available for displacement measurement rely on the changes in resistance, capacitance, or inductance resulting from the movement of a pendulum or cantilever. For achieving high sensitivity, a large proof mass is required. The sensor is therefore bulky and requires large power consumption. The electron tunneling accelerometer provides a high sensitivity along with a reduced sensor mass and lower power consumption.

4.13.2 Construction and Working

The accelerometer (Figure 4.23) consists of a thin glass plate covered by a 100 nm thick thermally evaporated Au film, which forms the Au electrode (Waltman and Kaiser 1989). The glass plate is fixed below a piezoelectric bimorph cantilever actuator; "bimorph" means an assembly of two piezo-electric crystals cemented together so that an applied voltage causes one to expand and the other to contract, converting electrical signals into mechanical energy. The Au electrode is positioned over an Au wire tunnel tip. The accelerometer is housed in an aluminum structure.

The Au electrode and the tunnel tip are suitably biased to start the tunneling current. The accelerometer functions with tunneling voltage and current over the ranges 0.01–1 V and $(0.05–10) \times 10^{-9}$ A. Coarse adjustment (lacking in delicacy or refinement) of the tunnel tip is done by a machine screw. The accelerometer operates through feedback control by comparing the measured tunneling current to a set point and maintaining a constant separation between the Au electrode and tunnel tip, thereby keeping the tunneling current constant. The accelerometer has a sensitivity of $10\,\mu g\ Hz^{-1/2}$ (where $g = 9.8\,m\ s^{-2}$) and a bandwidth of 3 kHz.

Example 4.4

The tunneling current I flowing between a pair of planar electrodes separated by a rectangular energy barrier of height ϕ and width s is given by

$$I = V \exp\left(-\alpha s \sqrt{\phi}\right) \qquad (4.77)$$

where $\alpha = 1.025\,eV^{-1/2}\ Å^{-1}$. The equation is valid for small bias voltages $V < \phi$.
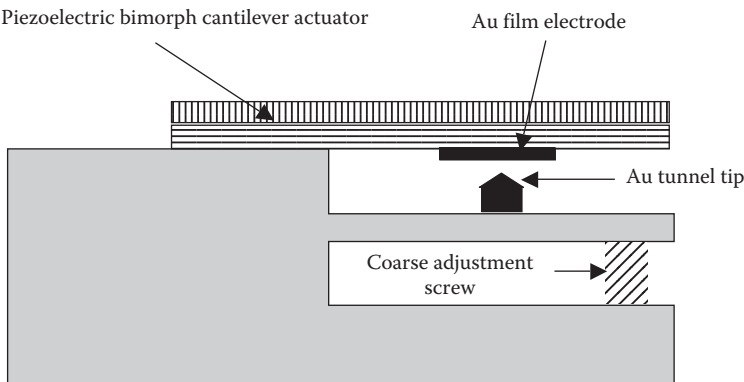
FIGURE 4.23
Silicon tunneling accelerometer. (After Waltman, S. B. and Kaiser, W. J., *Sensor. Actuator.*, 19, 201, 1989.)

Apply Equation 4.77 to find: (a) the factor by which the tunneling current increases for a decrease in separation between the electrodes by 1 Å for a typical barrier height $\phi = 3\,eV$; (b) the decrease in separation required to change the tunneling current by a factor of 10. (c) What is the value of separation for 10-fold change in tunneling current if the barrier height is 1 eV.

(a) When the separation decreases from s Å to $(s-1)$ Å, the tunneling current I_1 is

$$I_1 = V \exp\left\{-\alpha(s-1)\sqrt{\phi}\right\} \tag{4.78}$$

so that the ratio I_1/I is

$$\frac{I_1}{I} = \frac{V\exp\left\{-\alpha(s-1)\sqrt{\phi}\right\}}{V\exp\left\{-\alpha(s)\sqrt{\phi}\right\}} = \exp\left\{-\alpha(s-1-s)\sqrt{\phi}\right\}$$

$$= \exp\left\{\alpha\sqrt{\phi}\right\} = \exp\left\{(1.025)\sqrt{3}\right\} = \exp\left\{(1.025)\times 1.732\right\} = 5.90 \tag{4.79}$$

(b) If the decrease in separation required to change the tunneling current by one order of magnitude is denoted by x Å

$$10 = \frac{V\exp\left\{-\alpha(s-x)\sqrt{\phi}\right\}}{V\exp\left\{-\alpha(s)\sqrt{\phi}\right\}} = \exp\left\{-\alpha(s-x-s)\sqrt{\phi}\right\} = \exp\left(\alpha x\sqrt{\phi}\right) \tag{4.80}$$

Taking natural logarithm of both sides,

$$\ln(10) = \alpha x \sqrt{\phi} \tag{4.81}$$

from which

$$x = \frac{\ln(10)}{\alpha\sqrt{\phi}} = \frac{\ln(10)}{1.025\sqrt{3}} = \frac{2.303}{1.7753} = 1.297 \approx 1.3 \text{ Å} \tag{4.82}$$

(c)

$$x = \frac{\ln(10)}{\alpha\sqrt{\phi}} = \frac{\ln(10)}{1.025\sqrt{1}} = \frac{2.303}{1.025} \approx 2.25 \text{ Å} \tag{4.83}$$

4.13.3 Micromachined Accelerometer

Piezoelectric actuators are sensitive to temperature variations, hysteresis (the lagging of an effect behind its cause), and creep (the plastic deformation of a material under a constant load or force at elevated temperature) in the response of materials. Therefore, micromachined sensors employing electrostatic actuators were fabricated (Figure 4.24). These are immune to thermal drifts (drifts caused by internal heating of material during normal operation or by changes in external ambient temperature) and creep effects, and are easily miniaturized. A folded cantilever spring with integrated tip was resiliently (marked by the ability of bouncing back or recovering its shape, position, etc., readily after being stretched, bent, or compressed) mounted above a counter-electrode on a silicon substrate by means of a strip (beam) or rectangle attached at an end or edge to the substrate (Kenny et al. 1991, 1992, Rockstad et al. 1996, Dong et al. 2005). The strip or rectangle was produced by under-etching. Excursion of the resilient, cantilevered part caused by inertial forces arising from acceleration indicated the acceleration of the object. *Inertial force*, also known as effective force, is the fictitious force acting on a body as a result of using a non-inertial frame of reference, i.e., any force invoked by an observer to maintain the validity of Newton's second law of motion in a reference frame that is rotating
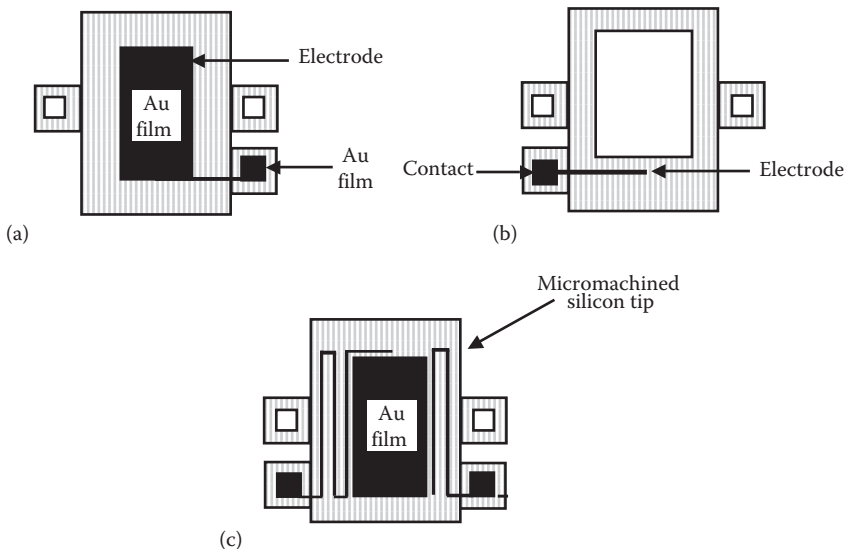
FIGURE 4.24
Three components of micromachined silicon tunnel sensor: (a) deflection counter electrode, (b) tunneling counter electrode, and (c) folded cantilever spring and tip. The tabs having square holes allow the components to be aligned properly. The inner rectangular area of part (c) can be deflected downward with respect to its outer segments by applying a voltage between the large electrode and corresponding deflection counterelectrode in part (a). (After Kenny, T. W. et al., *Appl. Phys. Lett.*, 58, 100, 1991.)
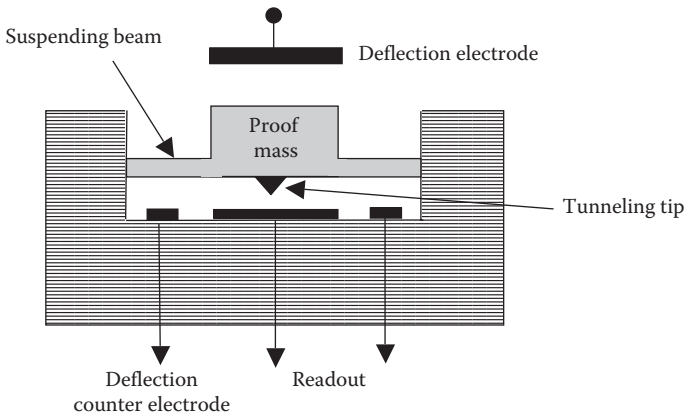
FIGURE 4.25
Tunneling accelerometer. (After Krishnan, G. et al., *J. Indian. I. Sci.*, 87, 333, 2007.)

or otherwise accelerating at a constant rate. Thus it is an imaginary force supposed to act upon an accelerated body, equal in magnitude and opposite in direction to the resultant of the real forces. Examples are the centrifugal and Coriolis forces that appear in rotating coordinate systems.

The accelerometer was constructed by assembling separate components, fabricated by lithographic techniques and having tabs with square holes that allowed the components to be laterally constrained by alignment pins. Figure 4.24 shows the components of the accelerometer. The displacement sensitivity was 10^{-4} Å $Hz^{-1/2}$. Figure 4.25 shows another version of the accelerometer (Krishnan et al. 2007).

Example 4.5

An accelerometer working on capacitive detection uses a cantilever-supported proof mass with a deflection sensitivity of 500 Å g^{-1} of acceleration. The capacitive signal of this accelerometer changes by 0.5% for a 1 g change of acceleration. Taking the barrier height as 3 eV, as in Example 4.4, find: (a) The deflection of the cantilever for a 0.5% change in tunneling current. (b) What is the improvement in signal response to acceleration in the tunneling accelerometer?

(a) $0.5\% = 0.5/100 = 0.005 = 5 \times 10^{-3}$. From Example 4.4, for an increase or decrease in tunneling current by a factor of 10, the required separation is 1.297 Å. Therefore, for change of tunneling current by a factor of 5×10^{-3}, it is $(1.297/10) \times 5 \times 10^{-3} = 6.485 \times 10^{-4}$ Å.

(b) 6.485×10^{-4} Å deflection of the tunneling accelerometer corresponds to 500 Å deflection of capacitive accelerometer. Hence, 1 Å deflection of the tunneling accelerometer = 500/ $(6.485 \times 10^{-4}) = 7.7 \times 10^5$ Å deflection of capacitive accelerometer. This means that the tunneling accelerometer is 7.7×10^5 times

superior than the capacitive one. In other words, the tunneling sensor technology results in an improvement of signal response by 7.7×10^5 times. The 6.485×10^{-4} Å deflection of tunneling accelerometer refers to an acceleration of $1/(7.7 \times 10^5) = 1.299 \times 10^{-6}$ g $= 1.299\,\mu g$.

4.14 NEMS Accelerometer

NEMS devices can be generally defined as devices fabricated by micro-technologies and including at least one movable mechanical part with dimension <100 nm. Ollier et al. (2006) designed two configurations: in-plane (IP) accelerometers with sensitive axis included in the plane of the wafer, and out-of-plane (OOP) accelerometers with sensitive axis perpendicular to the wafer (Figure 4.26). An acceleration causes the motion of a proof mass and the corresponding displacement is measured by means of a variation of capacitance between two sets of electrodes, which are either interdigitated combs (IP) or two parallel plates (OOP). In order to avoid parasitic deformations and undesirable vibration modes, a rigid mass was designed with typical dimensions of $75\,\mu m \times 150\,\mu m$. In addition, for IP accelerometers, low cross-sensitivity implies an increase of the ratio thickness/width of the mechanical springs, leading to 50 nm wide mechanical springs. In the same time, increase in electrical sensitivity requires a decrease in electrostatic gap to 100–400 nm. Typically, the length of springs and electrostatic teeth range from 1 to 10 μm. Some dimensions are now in the nanoscale, particularly the mechanical spring width (50 nm) and the electrostatic gaps (100–400 nm). The sensitivities are in the ranges of 100 and 500–1000 aF g^{-1} for IP and OOP accelerometers respectively; aF $= 10^{-18}$ F.
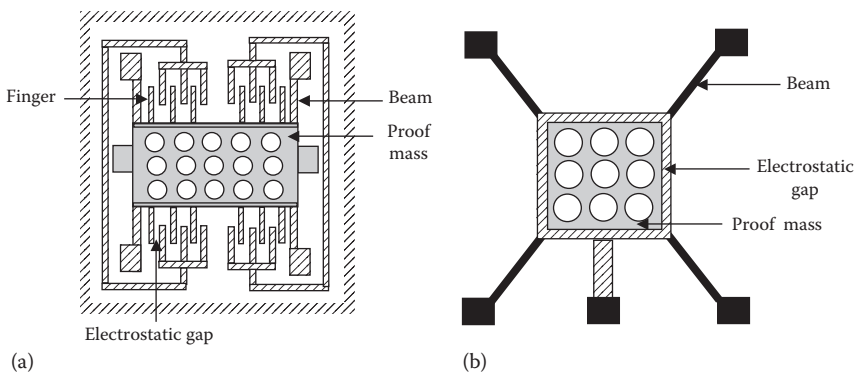
FIGURE 4.26
NEMS accelerometer: (a) in plane (IP) and (b) out-of-plane (OOP) configurations. (After Ollier, E. et al., *IEEE Sensors*, EXCO, Daegu, Korea, October 22–25, 54, 2006.)

4.15 Silicon Nanowire Accelerometer

Application of strain to a crystal causes a change in electrical conductivity due to the piezoresistance (PZR) effect. When a uniaxial stress is applied and the electric field and current are along the same direction, the longitudinal PZR coefficient can be measured. It is defined as the relative change in conductivity per unit stress. Si nanowires display an exceptionally large PZR effect compared with bulk silicon, e.g., the longitudinal PZR coefficient along the <111> direction increases with decreasing diameter for P-type Si nanowires, reaching as high as -3550×10^{-11} Pa^{-1}, in comparison with a bulk value of -94×10^{-11} Pa^{-1} (He and Yang 2006). This is the giant PZR effect. *What are the main features of this giant effect and what is its importance?* In giant PZR effect, the change in resistance of silicon nanowires due to an applied mechanical stress was reported to be orders of magnitude larger than that of bulk silicon (Milne et al. 2010). Additionally, giant PZR is seen as a potential breakthrough means of detecting motion in NEMS since PZR sensitivity scales better than optical or capacitive techniques. Moreover, since mechanical stress is a key element for performance enhancement of microelectronic devices, the physical mechanism behind giant PZR is vital for the design of accelerometers.

What is the interpretation of giant PZR effect? Still there is no consensus concerning the origins of giant PZR, although two models have some support: (i) One is based on a surface quantization effect predicted to occur in the first few silicon monolayers, while the (ii) other is based on a stress-induced shift of the surface Fermi level in depleted structures resulting from a change in surface charge. The atomic length scale of the former seems to be in disagreement with the typical wire diameters reported in the literature, which are at least several tens of nanometers, whereas the characteristic length scale of the latter is the surface depletion layer width (1–10 nm, depending on the doping density).

How is mechanical sensing performed using the giant PZR effect? Jointly using atomic-level simulation and experimental evaluation methods, Dao et al. (2010) elucidated the giant piezoresistive effect in single crystalline SiNWs along different crystallographic orientations. They applied SiNWs as nanoscale piezoresistors for mechanical sensing by constructing a Si NW accelerometer. In this accelerometer (Figure 4.27), the seismic mass was suspended on four surrounding sensing beams, which were themselves connected to the frame at the ends. Silicon nanowires were placed near the fixed ends on the surface of the sensing beams. Upon acceleration of the sensor, the seismic mass moved due to the inertial force. This motion deformed the beams so that the resistance of the nanowires was altered. By a Wheatstone bridge, the aforementioned change of resistance was converted to a voltage change. The overall size of the accelerometer chip was 500 μm × 500 μm × 400 μm. This accelerometer has been made from
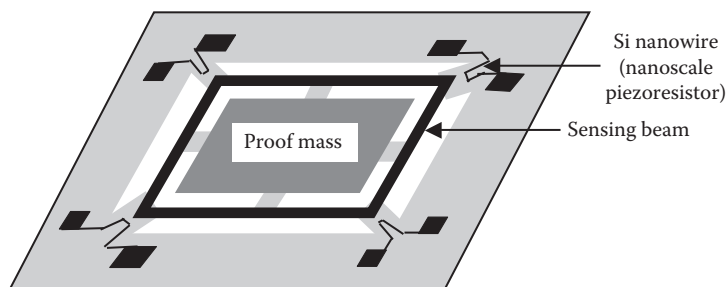
FIGURE 4.27
Silicon NW accelerometer. (After Dao, D. V. et al., *Adv. Nat. Sci.: Nanosci. Nanotechnol.*, 1: 013001, 10, 2010.)

multilayer SIMOX SOI wafer and fabricated by reactive ion silicon etching. The SiNW piezoresistor has width of 128 nm and thickness = 50 nm. The resistance of these piezoresistors was measured to be 20 kΩ. The calculated sensitivity for each axis was about 50 μV G^{-1}, and the resolution was 30 mG. Accordingly, by using SiNWs as piezoresistors, the sensor was miniaturized while increasing the sensitivity. Smaller chip size means larger number of chips per wafer, higher productivity, and therefore lower price per chip. Ultrasmall accelerometers are useful for portable devices, such as camcorders (video camera recorders), mobile phones, navigation systems, and entertainment devices.

4.16 CNT Flow Sensor for Ionic Solutions

Král and Shapiro (2001) showed theoretically the generation of an electric current in a metallic carbon nanotube immersed in a flowing liquid. They proposed a mechanism for the generation of the electric current and voltage through the transfer of momentum from the flowing liquid molecules to the acoustic phonons (quasiparticles representing the quantization of the modes of lattice vibrations of periodic, elastic crystal structures of solids) in the nanotube as the phonon quasi-momentum (a momentum-like vector). This in turn dragged free charge carriers in the nanotube producing the current flow.

Ghosh et al. (2003) reported that the flow of a liquid on SWCNT bundles induced a voltage in the sample along the direction of the flow. The voltage generated fitted logarithmic velocity dependence over nearly six decades of velocity. It was observed that the induced voltage tended to saturate at low flow velocities $\sim 10^{-5}$ m s^{-1}. Further, the magnitude of the voltage depended sensitively on the ionic conductivity and the polar nature of the liquid.

They put forward the notion of a flow sensor based on SWCNT in which an electrical signal was created in response to a fluid flow. They believed that this sensor could be scaled down to micrometer length dimensions comparable to the length of the individual nanotubes, making it usable for very small liquid volumes. The sensor had a high sensitivity at low velocities and a fast response time (better than 1 ms).

Bourlon et al. (2007) found that individual electrolytically gated SWCNT transistors of ~2 nm diameter, when incorporated into microfluidic channels, locally sensed the change in electrostatic potential induced by flow of ionic solutions on charged surfaces, known as the *streaming potential*. It is the potential which is produced when a liquid is forced to flow through a capillary tube or porous solid. It is the difference in electric potential at zero electric current between the capillary tube, porous solid, membrane, plug or diaphragm, and a liquid that is forced to flow through it. The streaming potential depends upon the presence of an electrical double layer at a solid–liquid interface. This electrical double layer is made up of ions of one charge type that are fixed to the surface of the solid and an equal number of mobile ions of the opposite charge which are distributed through the neighboring region of the liquid phase.

Taking help of the coupling between electrostatic potential and fluid dynamics, they demonstrated that individual nanotube transistors acted as fluidic flow sensors (Figure 4.28) that locally sensed potentials with nanometer-scale resolution and could be used to probe electrokinetic phenomena. These are the phenomena associated with movement of charged particles through a continuous medium or with the movement of a continuous medium over a charged surface. Examples are electrophoresis (the movement of colloidal particles or polyelectrolytes under an electric field), electroosmosis (the
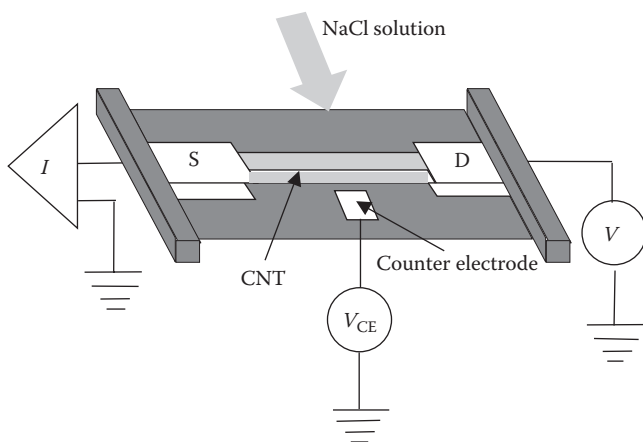
FIGURE 4.28
CNT-based flow sensor. (After Bourlon, B. et al., *Nat. Nanotechnol.*, 104, 104, 2007.)

motion of a liquid through an immobilized set of particles such as a capillary or porous plug, in response to an electric field), etc., involving tangential fluid motion adjacent to a charged surface and considered as manifestation of electrical properties of interfaces in steady-state and isothermal conditions.

How were the CNT transistors beneficial as flow sensors? Because of their ~2 nm height, nanotube devices were integrable into micro- or nanoscale fluidic circuits, making them appropriate for use in lab-on-a-chip applications, such as providing active flow sensing and high-resolution flow mapping on surfaces. Another benefit derived from this approach was that it enabled the study of nanotube conductance using the nanotube transconductance as an amplifier. Particularly, this ensured that the changes induced by the flow were generated locally from the nanotube only, getting rid of potential artifacts (spurious experimental results) that could originate at the electrode level.

They projected that the thermodynamic limit to sensitivity in a 1 Hz bandwidth was ~100 nL min^{-1} for a typical 100 μm × 200 μm microfluidic channel. To estimate the fundamental limits to flow sensing using this technique, they assumed that Johnson noise (the random fluctuations in the voltages due to the thermal agitation of charge carriers in a resistor; it is the mean-square EMF in conductors due to thermal agitation of the electromagnetic modes which are coupled to the thermal environment by the charge carriers) from the fluid resistance was the ultimate limit to the sensitivity. Johnson noise introduced thermal fluctuations δV_{CE} to the voltage applied to the electrolytic gate. They anticipated $\delta V_{CE} \sim 100$ mV Hz$^{-1/2}$ for $n = 10^{-6}$ M. Based on the measured values of α, the lowest detectable flow rate in the channel in a 1 Hz bandwidth was $Q = 100$ nL min^{-1}.

4.17 Discussion and Conclusions

Salient features of mechanical nanosensors are briefly presented in Table 4.4. Mass, displacement, acceleration, force, pressure, and fluid flow are the parameters that have been addressed for the fabrication of physical nanosensors. QCM has been long famous as an ultrasensitive acoustic wave–based mass measurement device. But MEMS and NEMS resonators have taken the lead and improved precision for measurement of infinitesimally small masses such as adsorbed atoms or molecules on surfaces and stretched the limits far below expectations.

Displacement sensing has progressed from electron tunneling devices to coulomb blockade electrometer and SET. Besides magnetomotive, piezo-based and optical options are available for displacement determination.

For measuring extremely small displacements and forces using CNTs, another approach has been pursued in which a bridge or cantilever presses

TABLE 4.4

Mechanical Nanosensors at a Glance

Sl. No.	Name of the Nanosensor	Sensed Quantity/ Applications	Nanomaterial/Nanostructure/ Microstructure Used	Sensing Principle
1	Quartz crystal microbalance (QCM)	Mass (up to 10^{-13} g)	Piezoelectric AT-cut quartz crystal	Decrease of acoustic wave resonance frequency with mass loading
2	MEMS/NEMS resonators	Mass (up to 10^{-18} and 10^{-21} g)	NEMS	Vibrational resonant frequency (GHz range) shift from adsorbed molecules
3	Electron tunneling displacement nanosensor	Displacement	Sharp tunneling tip	Variation of tunneling current with tunnel gap, as in STM
4	Coulomb blockade electrometer	Displacement	Two tunnel junctions with a capacitor coupling to the central region	Coulomb blockade effect
5	Single-electron transistor (SET)	Displacement	Capacitively coupled SET with a flexural beam resonator	Modulation of source–drain current by charge on gate electrode
6	Magnetomotive displacement nanosensor	Displacement	Conducting nanomechanical element	Induced EMF by a time-varying flux
7	Piezoresistive and piezoelectric displacement nanosensors	Displacement	NEMS	Resistance or current changes
8	Optical displacement nanosensor	Displacement	NEMS	Michelson and Fabry–Perot interferometry
9	Doubly clamped suspended CNT resonators	10^{-15} N force sensing	Carbon nanotube	Change in resonance frequency with applied force
10	Suspended CNT electromechanical sensors	Displacement and force	Suspended SWCNT fixed underneath a freestanding cantilever or bridge	Force acting on the cantilever or bridge deflects them and deforms the SWCNT below changing its conductance

(continued)

TABLE 4.4 (Continued)

Mechanical Nanosensors at a Glance

Sl. No.	Name of the Nanosensor	Sensed Quantity/ Applications	Nanomaterial/Nanostructure/ Microstructure Used	Sensing Principle
11	Membrane-based CNT pressure sensor	Pressure	CNT adhered to an alumina membrane	Piezoresistive effect
12	Tunnel effect accelerometer	Acceleration	MEMS	Relation of deflection of proof mass (measured through tunneling current changes) with acceleration
13	NEMS accelerometer	Acceleration	MEMS and NEMS	Capacitance variation between two sets of electrodes
14	Silicon nanowire accelerometer	Acceleration	Si NWs placed near the fixed ends on the surface of the sensing beams	Giant piezoresistive effect in single crystalline SiNWs
15	CNT flow sensor for ionic solutions	Liquid flow	CNT wall	Liquid flow produces a voltage along CNT wall depending on ionic conductivity of liquid; similar phenomenon occurs in electrolytically gated CNT transistors

downward on a CNT placed below it. The resulting deformation of the CNT causes a change in its resistance in accordance with the displacement or force through which the latter can be evaluated. Such kind of CNT-based displacement and force nanosensors have been developed.

For acceleration, MEMS accelerometer utilizing electron tunneling for displacement measurement, NEMS accelerometer and Si NW accelerometer have lowered significantly the detectable acceleration values. An ionic liquid flowing past a CNT creates a potential difference in relation to its flow rate which has been used in building liquid flow sensor for such samples.

Thus, several properties of nanomaterials or nanostructures have been advantageously put into service for measuring various mechanical quantities at hitherto unmatched accuracy levels.

Review Exercises

4.1 Describe the electrical actuation and detection of the guitar-string-like oscillation modes of doubly clamped nanotube oscillators. How can the devices be used to transduce very small forces?

4.2 Why is SiC an extremely promising material for high-frequency NEMS? Explain why for a given geometry, nanometer-scale SiC resonators are capable of yielding substantially higher frequencies than GaAs and Si resonators.

4.3 NEMS resonators have been suggested as nanosensors that allow the sensitive detection of molecules at atomic resolution. Explain.

4.4 SAW chemical sensors are based on the effects of adsorbed molecules on geometrical, elastic, and electric properties of gas-sensing layer and corresponding mass-loading of the working surface of the substrate carrying SAW. What is the SAW parameter which changes and is used for determining sensor output?

4.5 Explain how is the motion of a nanomechanical resonator detected by capacitively coupling the gate of the SET to a metal electrode placed on the resonator, and biasing the electrode at a constant voltage?

4.6 In a quartz crystal resonator, the crystal resonant frequency directly relates to the mass adsorbed; hence it is the parameter typically measured. Explain how the mass measurement is performed by inserting the crystal as the frequency controlling element of a feedback oscillator circuit.

4.7 When the resonance frequency shift Δf occurs due to mass change, mass change Δm is calculated using

$$\Delta m = -2\left(\frac{m}{f}\right)\Delta f$$

Applying this equation, prescribe the conditions regarding mass of the cantilever and its frequency to detect a small Δm.

4.8 Write the Sauerbrey equation between the frequency drop and mass loading of a QCM and explain the symbols used. State the conditions of validity of the equation.

4.9 Explain the operation of self-sensing cantilevers using piezoresistive displacement transduction.

4.10 Does the flow of a liquid on SWCNT bundles induce a voltage in the sample along the direction of the flow? How does the magnitude of the voltage depend sensitively on the ionic conductivity and on the polar nature of the liquid?

4.11 What happens when two conductors are brought into extreme proximity (~1 nm) with an applied bias between them? Name the phenomenon and show how it is utilized in the accelerometer.

4.12 What is Coulomb blockade effect? What is a Coulomb blockade electrometer? Compare the Coulomb blockade electrometer and electron tunneling transistor as displacement sensing devices.

4.13 How are individual CNT transistors useful for probing electrokinetic phenomena and fluid flow in microfluidic channels? Can these fluidic flow sensors locally sense potentials with nanometer-scale resolution?

4.14 For cantilever devices with sizes comparable to the mean free path in air, the added benefit of preservation of a high-quality factor in ambient enables unprecedented, sub-attogram-scale mass resolution at room temperature and atmospheric pressure. Elaborate this statement to explain that nanoscale mechanical sensors offer a greatly enhanced performance than is unattainable with microscale devices.

4.15 The Langmuir–Blodgett (LB) technique is a molecular engineering technique able to fabricate materials with high structural order and sensing multilayers, monolayer by monolayer, with fine thickness control. How are SWCNT films prepared by the LB technique on a sensor surface for volatile organic compound detection?

4.16 Mention a few advantages of electron tunneling displacement transducers. Draw and explain the operation of a micromachined electron tunneling displacement transducer.

4.17 Describe two techniques for actuation of NEMS devices. Explain also how these techniques are utilized for motion detection.

4.18 An important concept in the study of electromechanical transducers is backaction. Backaction describes the perturbation of the mechanical element by the sensor and the output circuit during the detection. Define backaction noise.

References

Albrecht, T. R., P. Grtitter, D. Horne, and D. Rugar. 1991. Frequency modulation detection using high-*Q* cantilevers for enhanced force microscope sensitivity. *Journal of Applied Physics* 69(2): 668–673.

Anantram, M. P. and F. L'eonard. 2006. Physics of carbon nanotube electronic devices. *Reports on Progress in Physics* 69: 507–561.

Ansari, M. Z., C. Cho, J. Kim, and B. Bang. 2009. Comparison between deflection and vibration characteristics of rectangular and trapezoidal profile microcantilevers. *Sensors* 9: 2706–2718.

Beeby, S., G. Ensell, M. Kraft, and N. White. 2004. *MEMS Mechanical Sensors*. Boston, MA: Artech House, Inc., 269 pp.

Bourlon, B., J. Wong, C. Miko, L. S. Forro, and M. Bockrath. 2007. A nanoscale probe for fluidic and ionic transport. *Nature Nanotechnology* 104(2): 104–107.

Chen, G. Y., T. Thundat, E. A. Wachter, and R. J. Warmack. 1995. Adsorption-induced surface stress and its effects on resonance frequency of microcantilevers. *Journal of Applied Physics* 77(8): 3618–3622.

Cleland, A. N. and M. L. Roukes. 1996. Fabrication of high frequency nanometer scale mechanical resonators from bulk Si crystals. *Applied Physics Letters* 69: 2653–2655.

Dai, M. D., K. Eom, and C.-W. Kim. 2009. Nanomechanical mass detection using nonlinear oscillations. *Applied Physics Letters* 95: 203104.

Dao, D. V., K. Nakamura, T. T. Bui, and S. Sugiyama. 2010. Micro/nano-mechanical sensors and actuators based on SOI-MEMS technology. *Advances in Natural Sciences: Nanoscience and Nanotechnology* 1: 013001 (10pp).

Dong, H., Y. Jia, Y. Hao, and S. Shen. 2005. A novel out-of-plane MEMS tunneling accelerometer. *Sensors and Actuators A* 120: 360–364.

Duemling, M. 2002. *Modeling and Characterization of Nanoelectromechanical Systems*. MS Thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA, pp. 6–13.

Ekinci, K. L. 2005. Electromechanical transducers at the nanoscale: Actuation and sensing of motion in nanoelectromechanical systems (NEMS). *Small* 1(8–9): 786–797.

Ekinci, K. L., Y. T. Yang, and M. L. Roukes. 2004. Ultimate limits to inertial mass sensing based upon nanoelectromechanical systems. *Journal of Applied Physics* 95(5): 2682–2689.

Ghosh, S., A. K. Sood, and N. Kumar. 2003. Carbon nanotube flow sensors. *Science* 299: 1042–1044.

Hansen, K. M. and T. Thundat. 2005. Microcantilever biosensors. *Methods* 37: 57–64.

He, R. and P. Yang. 2006. Giant piezoresistance effect in silicon nanowires. *Nature Nanotechnology* 1: 42–46, doi: 10.1038/nnano.2006.53

Ho, C. K., E. R. Lindgren, K. S. Rawlinson, L. K. McGrath, and J. L. Wright. 2003. Development of a surface acoustic wave sensor for *in-situ* monitoring of volatile organic compounds. *Sensors* 3(7): 236–247, doi: 10.3390/s30700236

Huang, X. M. H., X. L. Feng, C. A. Zorman, M. Mehregany, and M. L. Roukes. 2005. VHF, UHF and microwave frequency nanomechanical resonators. *New Journal of Physics* 7(247): 1–15, doi: 10.1088/1367-2630/7/1/247.

Huang, X. M. H., C. A. Zorman, M. Mehregany, and M. L. Roukes. 2003. Nanoelectromechanical systems: Nanodevice motion at microwave frequencies. *Nature* 421: 496.

Ikehara, T., J. Lu, M. Konno, R. Maeda, and T. Mihara. 2007. A high quality-factor silicon cantilever for a low detection-limit resonant mass sensor operated in air. *Journal of Micromechanics and Microengineering* 17: 2491–2494.

Jenkins, N. E., L. P. DeFlores, J. Allen, T. N. Ng, S. R. Garner, S. Kuehn, J. M. Dawlaty, and J. A. Marohn. 2004. Batch fabrication and characterization of ultrasensitive cantilevers with submicron magnetic tips. *Journal of Vacuum Science and Technology B* 22(3): 909–915.

Kanazawa, K. and N.-J. Cho. 2009. Quartz crystal microbalance as a sensor to characterize macromolecular assembly dynamics. *Journal of Sensors* 2009, Article ID 824947, 17 pp., doi:10.1155/2009/824947

Kenny, T. W., W. J. Kaiser, J. K. Reynolds, J. A. Podosek, H. K. Rockstad, E. C. Vote, and S. B. Waltman. 1992. Electron tunnel sensors. *Journal of Vacuum Science Technology A* 10(4): 2114–2118.

Kenny, T. W., S. B. Waltman, J. K. Reynolds, and W. J. Kaiser. 1991. Micromachined silicon tunnel sensor for motion detection. *Applied Physics Letters* 58(1): 100–102.

Knobel, R. and A. N. Cleland. 2002. Piezoelectric displacement sensing with a single-electron transistor. *Applied Physics Letters* 81(12): 2258–2260.

Knobel, R. G. and A. N. Cleland. 2003. Nanometre-scale displacement sensing using a single electron transistor. *Nature* 424: 291–293.

Král, P. and M. Shapiro. 2001. Nanotube electron drag in flowing liquids. *Physical Review Letters* 86: 131–134.

Krishnan, G., C. U. Kshirsagar, G. K. Ananthasuresh, and N. Bhat. 2007. Micromachined high-resolution accelerometers. *Journal of the Indian Institute of Science* 87(3): 333–361.

Lavrik, N. V., M. J. Sepaniak, and P. G. Datskos. 2004. Cantilever transducers as a platform for chemical and biological sensors. *Review of Scientific Instruments* 75 (7): 2229–2253.

Li, C. and T.-W. Chou. 2003. Single-walled carbon nanotubes as ultrahigh frequency nanomechanical resonators. *Physical Review* B68L:073405-1–073405-3.

Li, C.-Y. and T.-W. Chou. 2004. Strain and pressure sensing using single-walled carbon nanotubes. *Nanotechnology* 15: 1493–1496.

Li, C. and T.-W. Chou. 2006. Atomistic modeling of carbon nanotube-based mechanical sensors. *Journal of Intelligent Material Systems and Structures* 17: 247–254.

Li, X., T. Ono, Y. Wang, and M. Esashi. 2003. Ultrathin single-crystalline-silicon cantilever resonators: Fabrication technology and significant specimen size effect on Young's modulus. *Applied Physics Letters* 83(15): 3081–3083.

Lin, Y.-C., T. Ono, and M. Esashi. 2005. Fabrication and characterization of microma-chined quartz-crystal cantilever for force sensing. *Journal of Micromechanics and Microengineering* 15: 2426–2432.

Lu, J., T. Ikehara, Y. Zhan, T. Mihara, T. Itoh, and R. Maeda. 2007. High quality factor silicon cantilever transduced by piezoelectric lead zirconate titanate film for mass sensing applications. *Japanese Journal of Applied Physics* 46(12): 7643–7647.

Milne, J. S., A. C. H. Rowe, S. Arscott, and Ch. Renner. 2010. Giant piezoresistance effects in silicon nanowires and microwires. *Physical Review Letters* 105(22): 226802.

Minot, E. D., Y. Yaish, V. Sazonova, J.-Y. Park, M. Brink, and P. L. McEuen. 2003. Tuning carbon nanotube band gaps with strain. *Physical Review Letters* 90(15): 156401–156404.

Ollier, E., L. Duraffourg, M. Delaye, S. Deneuville, V. Nguyen, P. Andreucci, H. Grange et al. 2006. Thin SOI NEMS accelerometers compatible with in-IC integration. In: *IEEE Sensors* 2006, EXCO, Daegu, Korea, October 22–25, 2006, pp. 54–57.

Rast, S., C. Wattinger, U. Gysin, and E. Meyer. 2000. The noise of cantilevers. *Nanotechnology* 11: 169–172.

Rinaldi, G. 2009. Simple and versatile micro-cantilever sensors. *Sensor Review* 29(1): 44–53.

Rockstad, H. K., T. K. Tang, K. Reynolds, T. W. Kenny, W. J. Kaiser, and T. B. Gabrielson. 1996. A miniature, high-sensitivity, electron tunneling accelerometer. *Sensors and Actuators A* 53: 227–231.

Sazonova, V., Y. Yaish, H. Ustunel, D. Roundy, T. A. Arias, and P. L. McEuen. 2004. A tunable carbon nanotube electromechanical oscillator. *Nature* 431: 284–287.

Sone, H., Y. Fujinuma, and S. Hosaka. 2004. Picogram mass sensor using reso-nance frequency shift of cantilever. *Japanese Journal of Applied Physics* 43(6A): 3648–3651.

Stampfer, C., T. Helbling, D. Obergfell, B. Schöberle, M. K. Tripp, A. Jungen, S. Roth, V. M. Bright, and C. Hierold. 2006a. Fabrication of single-walled carbon-nano-tube-based pressure sensors. *Nano Letters* 6(2): 233–237.

Stampfer, C., A. Jungen, and C. Hierold. 2005. Nano electromechanical trans-ducer based on single-walled carbon nanotubes. In: *Transducers 05. The 13th International Conference on Solid-State Sensors, Actuators and Microsystems*, Seoul, Korea, June 5–9, 2005, pp. 2103–2106.

Stampfer, C., A. Jungen, and C. Hierold. 2006b. Fabrication of discrete nanoscaled force sensors based on single-walled carbon nanotubes. *IEEE Sensors Journal* 6: 613–617.

Stampfer, C., A. Jungen, R. Linderman, D. Obergfell, S. Roth, and C. Hierold. 2006c. Nano-electromechanical displacement sensing based on single-walled carbon nanotubes. *Nano Letters* 6(7): 1449–1453.

Stampfer, C., A. Jungen, R. Linderman, D. Obergfell, S. Roth, and C. Hierold. 2006d. Nano-electromechanical displacement sensing based on single-walled carbon nanotubes. *Nano Letters* 6: 1449–1453.

Vashist, S. K. 2007. A review of microcantilevers for sensing applications. *Journal of Nanotechnology Online* 3: 1–15.

Waltman, S. B. and W. J. Kaiser. 1989. An electron tunneling sensor. *Sensors and Actuators* 19(3): 201–210.

Wang, K., A.-C. Wong, and C. T.-C. Nguyen. 2000. VHF free–free beam high-micro-mechanical resonators. *Journal of Microelectromechanical Systems* 9(3): 347–360.

Wautelet, M. 2001. Scaling laws in the macro-, micro- and nanoworlds. *European Journal of Physics* 22: 601–611.

White, J. D. 1993. An ultra high resolution displacement transducer using the Coulomb blockade electrometer. *Japanese Journal of Applied Physics* 32: L1571–L1573.

Yang, L., M. P. Anantram, J. Han, and J. P. Lu. 1999. Band-gap change of carbon nano-tubes: Effect of small uniaxial and torsional strain. *Physical Review B* 60(19): 13874–13878.

Yang, J., T. Ono, and M. Esashi. 2000. Mechanical behavior of ultrathin microcantile-ver. *Sensors and Actuators* 82: 102–107.

5

Thermal Nanosensors

5.1 Introduction

Temperature is a fundamental physical property in both scientific and industrial fields. It is one of the primary physical quantities that is routinely measured to derive other thermodynamic quantities such as heat, energy, or specific heat capacity (the ratio of the amount of heat, measured in calories, required to raise the temperature of 1 g of a substance by 1°C to the amount of heat required to raise the temperature of a similar mass of a reference material, usually water, by the same amount); *thermodynamics* is the science that deals with the relationships and conversions between heat and other forms of energy, most notably mechanical work. Temperature sensing is therefore a vital requirement in many parts of industry including aerospace (comprises the atmosphere of the Earth and surrounding space, typically referring to industry concerned with vehicles moving through air and space), nuclear, mechanical, chemical, and medical technologies. If these sensors are small in size, they can be readily embedded in the systems without intrusion.

As semiconductor devices continue to decrease in size, and more and more applications are found for the micro- and nanodevices, there is an urgent need for measuring physical quantities at submicrometer or even nanometer-size dimensions. The burgeoning field of nanotechnology promises to revolutionize many scientific fields.

What are the existing methods for temperature measurement? Presently, temperature is measured using thermocouples (TCs) (temperature sensors consisting of two dissimilar metallic wires, e.g., one chromel and one constantan wire), semiconductor diodes working on the dependence of their forward voltages on temperature, resistance temperature detectors (RTD, commonly made of platinum), thermistors (temperature-sensitive resistors made from barium titanate ($BaTiO_3$) or the oxides of the transition metals: manganese (Mn), cobalt (Co), copper (Cu), and nickel (Ni)), and infrared (wavelength = 0.7–300 μm) thermometers (noncontact thermometers determining temperature from the blackbody radiation of a body; a black body is an idealized physical body that absorbs all incident electromagnetic radiation). Although infrared radiation is invisible, humans can sense it as heat. Astronomers have found that infrared radiation is particularly useful when trying to probe areas of

our universe that are surrounded by clouds of gas and dust. Because of longer wavelength of infrared radiation, it can pass right through these clouds and reveal details invisible by observing other types of radiation. Especially interesting are areas where stars and planets are forming and the cores of galaxies where it is believed that huge black holes reside.

Further ahead are near-field optical thermometers, and other exotic methods. The spatial resolution of infrared thermometry is limited to the wavelength of collected IR radiation due to diffraction effects (the apparent bending of waves around small obstacles and the spreading out of waves past small openings). Temperature of objects inaccessible to far-field infrared thermometry is found by near-field thermometry. For ultrahigh optical resolution, near-field scanning optical microscopy (NSOM) is currently the photonic instrument of choice.

The near-field is defined as the region above a surface with dimensions less than a single wavelength of light incident on the surface. Near-field imaging takes place when a submicron optical probe is positioned at a very short distance from the sample and light is transmitted through a small aperture at the tip of this probe. Within the near-field region, evanescent light is not diffraction limited and nanometer spatial resolution is achievable. Evanescent is defined as "temporal" or "fleeting"; unlike normal light, evanescent light is extremely limited in terms of the distance it can be propagated.

Lateral resolution of 20 nm and vertical resolution of 2–5 nm have been demonstrated. The collected radiation depends on the temperature-dependent thermal reflectance (the fraction of incident radiation reflected by a surface) of the sample.

Each of the techniques mentioned earlier has its own advantages and disadvantages, e.g., the TC is extremely simple in design and inexpensive while the semiconductor diode is unsuitable for high temperatures but used extensively to measure cryogenic temperatures. The spatial resolution of most of these techniques is about 10 μm, except for the near-field thermometry in which the spatial resolution can be in the order of 50 nm. But near-field thermometry involves complicated optical instrumentation and is quite a cumbersome technique.

5.2 Nanoscale Thermocouple Formed by Tungsten and Platinum Nanosize Strips

How is a nanosized thermocouple junction made? Fabrication of a nanosized sensor (Figure 5.1) (Chow et al. 2002) consists of the following two steps: (i) Deposition of a nanosized strip of first metal (tungsten) on an electrical insulator substrate by focused ion beam (FIB) deposition process. (ii) Deposition of a nanosized strip of second metal (Platinum) on the same
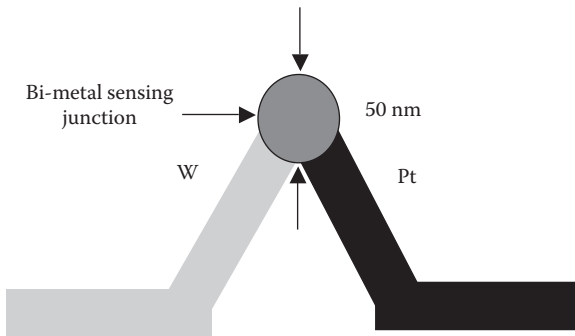
FIGURE 5.1
Overlapped nanosensor structure. (After Chow, L. et al. Fabrication of nano-scale temperature sensors and heaters. United States Patent 6905736, Filing Date: 02/27/2002, Publication Date: 06/14/2005.)

substrate by FIB process in a partially overlapping portion on first metal nanosized strip.

The partially overlapping segment is selected from one of a ball-shaped portion, and a point-shaped configuration portion, wherein the first metal nanosized strip and the second metal nanosized strip both have a thickness around 50 nm, and constitute a bimetal sensing junction from the partially overlapping portion between the first and the second strips. The cross-sectional area of the bimetal sensing junction is approximated as $50 \times 50\,nm^2$.

The process used to deposit metal nano strips to fabricate the sensor is ion-beam-assisted chemical vapor deposition (CVD). The FIB Ga^+ ion beam is used to decompose $W(CO)_6$ molecules to deposit a tungsten nanostrip on a suitable substrate. For the Pt strip, the precursors are trimethyl platinum $(CH_3)_3Pt$. Because of the use of Ga^+ beam in the deposition, both Pt and W nanostrips contain a definite proportion of Ga impurities, which are denoted as Pt(Ga) and W(Ga), respectively. By characterization of the response of this Pt(Ga)/W(Ga) nanoscale junction, a temperature coefficient of ~5.4 mV °C^{-1} was specified. This was a factor of ~130 larger than the conventional K-type TCs. K-type TC (which is made of a nickel–chromium/nickel–aluminum junction, called chromel/alumel) is the "general purpose" TC. K-type TCs are available in the 95°C–1260°C range. Sensitivity is approx 41 µV °C^{-1}.

5.3 Resistive Thermal Nanosensor Fabricated by Focused Ion Beam Chemical Vapor Deposition

Can a resistive nanosensor be made by FIB method? Focused ion beam chemical vapor deposition (FIB-CVD) is a very useful technique to make three-dimensional nanostructures and can be applied to fabricate various

nanodevices such as nanosensors. Ozasal et al. (2004) produced tungsten thermal nanosensor with 3D nanostructure by FIB-CVD using $W(CO)_6$ source gas material. Four-terminal Au electrodes were formed on a 100 nm thick SiO_2 on Si substrate by electron-beam lithography and dry etching. Then, a thermal nanosensor with 100 nm diameter was made by FIB-CVD. The exposure time was 9 min.

The temperature measurement using nanosensor was based on the change of electrical resistivity with temperature. The resistivity of tungsten deposited by FBI-CVD was about 100 times higher than that of bulk tungsten at 25°C. The resistivity of thermal nanosensor can be correctly measured by four-terminal electrode arrangement as the contact resistances are ignored in this method. Four-point probe method is an electrical impedance measuring technique that uses separate pairs of current-carrying and voltage-sensing electrodes to carry out more accurate measurements than traditional two-terminal (2T) sensing.

The results indicated that the resistivity of thermal nanosensor was proportional to temperature from room temperature to 120°C.

5.4 "Carbon-Nanowire-on-Diamond" Resistive Temperature Nanosensor

A temperature sensor based on an array of carbon nanowires written by a 30 keV Ga[+] FIB on diamond substrate was developed by Zaitsev et al. (2007). *What is the reason for using carbon nanowires in place of carbon nanotubes (CNTs)?* The use of carbon nanowires in NEMS and nanoelectronics is considerably facilitated by their far more flexible technology compared to CNTs, controllability of their positioning and reproducibility of their electrical and structural parameters.

The sensor (Figure 5.2) showed an exponential increase of current with temperature at a rate of 0.1 dB °C[−1] in the temperature range from 40°C to 140°C. The conductance along nanowires was high and ohmic, while the conductance in the orthogonal direction was much lower and superlinear. The advantages of the novel carbon nanowire sensor are light blindness, compatibility with carbon nanotechnology, reproducibility, and simplicity of fabrication.

What is the explanation for the temperature-dependent characteristic of a nanowire? The temperature sensitivity of the electrical conductance between the nanowires is interpreted by the thermal activation of charge carriers in heterostructure graphite-diamond over an energy barrier of 0.24 eV; a heterostructure is a structure of two different materials.
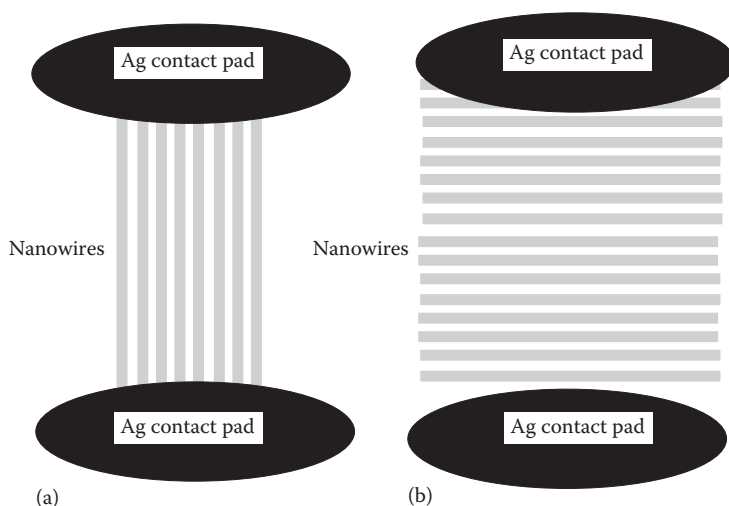
FIGURE 5.2

Carbon nanowire temperature sensor: (a) nanowires fixed parallel to contact pads and (b) nanowires fixed perpendicular to contact pads. (After Zaitsev, A. M. et al., *Phys. Status Solidi A*, 204, 3574, 2007.)

5.5 Carbon Nanotube Grown on Nickel Film as Resistive Low-Temperature (10–300 K) Nanosensor

Saraiya et al. (2006) grew CNTs on nickel film deposited on float glass substrate using an ion beam deposition technique (Figure 5.3), a process of applying thin films of materials on a surface through the application of an
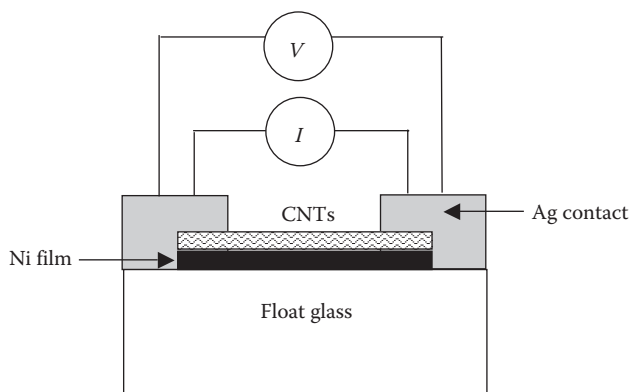
FIGURE 5.3

CNT low-temperature sensor. (After Saraiya, A. et al., *Syn. React. Inorg. Met.*, 36, 163, 2006.)

ion beam, either directly from an ion source or by sputtering from a target bombarded by ion beam. The percentage change of resistance of CNT between 300 and 10 K was more than 600%. The drastic change in resistance at low temperature indicates that CNTs grown on nickel film can be used as low-temperature sensors.

5.6 Laterally Grown CNT between Two Microelectrodes as Resistive Temperature Nanosensor

The obvious question is: *What is the reason for insistence on using CNTs? How is the use of CNT as a temperature-sensing element beneficial?* In a complicated thermal flow system, the perturbation to the system from the temperature sensor becomes critically important and cannot be neglected. The advantage derived by using CNT as a sensor is that the extremely small size of the CNT provides accurate measurements at nanoscale size without disturbing the neighboring environment. In addition, the power consumption is very low for a small-size sensor. Also, CNT can provide an extremely small time constant giving extremely rapid time response with the temperature of the object on which measurement is done.

Kuo et al. (2007) presented direct growth or fabrication of CNTs between MEMS-electrodes. A high-quality CNT sensor (Figure 5.4) was obtained by adopting properly controlled growth conditions. After a thin oxide was deposited on silicon wafer, a thin film of nickel was formed and used to make nanoparticles by sputter etching, a process wherein sputtering is used to remove atoms from a material. These nanoparticles were used as catalysts to grow CNT at a later stage. Then, titanium film was deposited and patterned as electrodes and metal leads. When the titanium film was patterned,
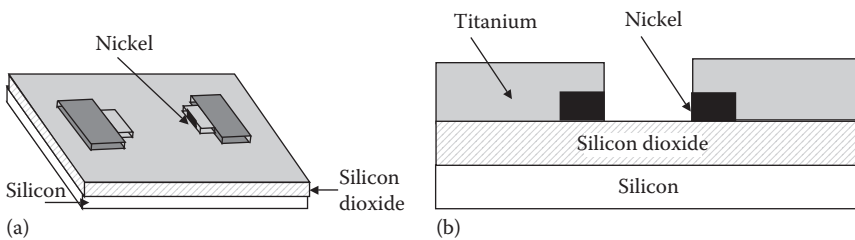
FIGURE 5.4
Electrode arrangement of temperature nanosensor chip: (a) top view and (b) cross section. CNT (not shown here) is grown laterally between electrodes. (After Kuo, C. Y. et al., *IEEE Trans. Nanotechnol.*, 6, 63, 2007.)

the nanoparticles were exposed on the sidewall of the electrodes and these were used to grow CNT in the horizontal direction.

The above workers grew CNT laterally and selectively between electrodes made by MEMS techniques. The CNT growth method employed was microwave plasma chemical vapor deposition (MPCVD). After electrodes were made, the silicon chip was transferred into an electron cyclotron resonance chemical vapor deposition (ECR-CVD) chamber, which was powered by microwaves, to grow CNTs. ECR is the resonance absorption of energy from a radio-frequency or microwave-frequency electromagnetic field by electrons in a uniform magnetic field when the frequency of the electromagnetic field equals the cyclotron frequency of the electrons. In an ECR source, vapor of the desired element is held in a specially designed magnetic field long enough for the elemental atoms to be ionized in collisions with electrons, which are kept in motion by microwaves.

The nanotubes grown were actually multi-wall nanotubes. The source gas used was CH_4, which was varied from 10% to 60% during CNT growth process. The addition of N_2 gas increased the bonding force among carbon atoms and enhanced the carbon activity when they precipitated from the catalyst. Thus, better-quality CNTs could be made with a high degree of graphitization (the formation of graphite-like material from organic compounds). Therefore, a different amount of N_2 gas was fed into the chamber to obtain a superior quality of graphitization and to make CNTs grow straight, like their growth in the vertical direction. Other growth parameters were chamber pressure = 30 torr, RF power of the plasma = 800 W, and the substrate temperature = 400°C. Both SEM and Raman spectrum analysis were used to determine a better set of growth parameters in the MPCVD process. The amount of methane and nitrogen gas had to be properly controlled to obtain a successful better-quality CNT, connecting the two neighboring electrodes.

After growth of CNTs, the silicon chip was wire bonded on a polycarbonate (PC) board. The electric signals acquired at the CNT were fed to a computer for analysis and plotting. The results indicated the very linear variation of the resistance with the temperature (20°C–160°C). This suggested that the laterally grown CNT can be used as temperature sensor despite the seriously curled structure. However, two different types of CNTs were distinguished: one which showed increase in resistance with the temperature indicated the metallic nature, while the other that showed decrease in the resistance with increasing temperature had semiconducting nature.

It is noted that the CNT grown may be metallic or semiconductor due to the changes in the carbon network structure. However, the multi-wall nature of CNTs grown by MPCVD suggests that some of the layers may be metallic, but others may be semiconducting. This makes the sensor become metallic since the semiconductor part may not conduct electricity, or it may have a very high resistance due to the amount of dopants

absorbed. To become a semiconducting sensor, all of the wall layers in the CNT must be semiconductors. The probability that all of the layers in the CNT become semiconductors is very low. Therefore, one does not often obtain a semiconducting sensor from a multi-walled carbon nanotube (MWCNT).

What is I–V behavior and TCR of the CNTs? The linear *I–V* curve of CNT indicates that the titanium/MWCNT junction is an ohmic contact. The linear relationship of the electrical resistance with the temperature and the temperature response or recovery test of the CNT showed that the CNT grown laterally between electrodes can be used as a good temperature sensor despite the fact that the CNT is grown either in metallic or semiconductor nature.

For metallic CNT, the resistance is (Kuo et al. 2007)

$$R = 1.1956T + 1834.2 \tag{5.1}$$

For semiconducting CNT, the resistance is

$$R = -0.8976T + 5344.3 \tag{5.2}$$

where T is the temperature in °C. The temperature coefficient of resistance (TCR) for the current MWCNT sensor, which was grown at $CH_4 = 20\%$ and $N_2 = 20$ sccm, obtained within the experimental growth conditions and gap widths, varied from 0.0008152 to 0.0001759. The TCR is quite low as compared with the TCR using platinum as temperature sensor. However, the nanosize of CNT can have a very high response and sensitivity to the environmental temperature change. This is highly useful for measurements in systems with very rapid temperature variations.

Example 5.1

Compare the resistances of metallic and semiconducting CNTs at 100°C and 150°C. Is there any difference in the nature of variation?

For metallic CNT, from Equation 5.1, at 100°C

$$R_1 = 1.1956T + 1834.2 = 1.1956 \times 100 + 1834.2 = 1953.76 \ \Omega \tag{5.3}$$

Similarly at 150°C, $R_1 = 2013.54 \,\Omega$.

For semiconducting CNT, from Equation 5.2,

$$R_2 = -0.8976T + 5344.3 = -0.8976 \times 100 + 5344.3 = 5254.54 \ \Omega \tag{5.4}$$

Likewise, at 150°C, $R_2 = 5209.66 \,\Omega$.

R_1 increases with temperature whereas R_2 decreases with temperature. This is expected in conformance with the metallic and semiconducting behavior of the CNTs.

5.7 Silicon Nanowire Temperature Nanosensors: Resistors and Diode Structures

The silicon nanowires (SiNWs) are accredited as temperature sensors because of the fact that the conductance of these semiconductor wires is highly dependent on temperature. Agarwal et al. (2008) presented SiNWs as nano temperature sensors. They employed two configurations, namely, (i) resistance temperature detector (RTD) and (ii) diode temperature detector (DTD) types (Figure 5.5). The first configuration of RTD-type sensors showed TCR values ~7500 ppm K^{-1}, which were enhanced beyond 10,000 ppm K^{-1} by the application of back-bias. The second configuration of DTD-type sensors using nanowires recorded more than one order variation in reverse-bias current, in the temperature range of 293–373 K. Both types of nano temperature sensors were highly sensitive and could be integrated with other biochemical sensors in lab-on-a-chip devices that integrate one or several laboratory functions on a single chip of only millimeters to a few square centimeters in size.

What advantages favor SiNWs for temperature sensing? The SiNW approach paved the way for realizing high-density, highly sensitive biochemical sensor chips integrated with nano temperature sensors using mass fabrication technologies and thus had the potential to achieve multiplexed diagnosis employing nanowire arrays (Figure 5.6), which was essential for understanding and curing of complex diseases. The approach also had the inherent advantage of integrating with Si-based signal processing and communication circuits. Conversely, most present techniques reported for nano temperature sensors, like assembly of CNT as thermal probe, face mass fabrication challenges. The mass manufacturing capability, reproducibility, and integration with other functionalities in a biochip is extremely challenging for these sensors.

How are Si nanowire sensors fabricated? SiNWs were realized on 8″ diameter wafers. For single-crystal SiNWs, silicon-on-insulator (SOI) wafers were used. For polysilicon nanowires, 80 nm thick polysilicon was deposited on 50 nm thick thermal oxide grown on P-type test wafer. The wafers were doped with P-type (boron) or N-type (phosphorous) impurities by ion implantation. The dopants were activated in a rapid thermal annealing (RTA) furnace (short-time heat treatment furnace) and nanowire-fins were patterned using deep ultraviolet lithography in the array format. Generally, techniques for RTA
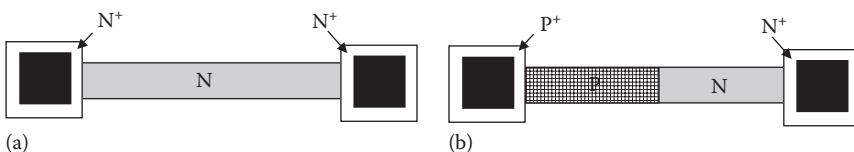
FIGURE 5.5
(a) Nanowire resistor and (b) nanowire diode. (After Agarwal, A. et al., *Sensor Actuat.* A, 145–146, 207, 2008.)
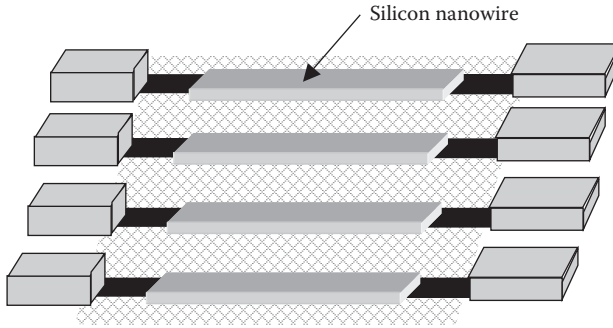
FIGURE 5.6
Nanowire array with metal lines and contacts to individual nanowires. (After Agarwal, A. et al., *Sensor Actuat. A,* 145–146, 207, 2008.)

fall into two main types depending on the heating method: furnaces using steady heat sources or those employing electrical lamps, either incandescent or arc, with programmed optical output cycles.

The two ends of the nanowires were further doped to obtain N^+ or P^+ regions, followed by connecting to contact metal and alloying to realize ohmic contacts. The device was passivated by silicon nitride film except for the active nanowire sensor area and metal pads.

How does the device measure temperature? Temperature–conductance $\sigma(T)$ curve for several polysilicon nanowires showed the linear increase in conductance with temperature having a standard deviation of ~85 pS over the set of 10 wires at 20°C. By suitable back-gate bias (V_G) applied from the back of the Si substrate, the absolute value of TCR could be increased to more than 10,000 (at $V_G = 4$ V). The relative change in conductance with temperature was appreciably more for polysilicon nanowires than single-crystal nanowires.

How are DTDs fabricated? The process of fabricating a DTD is identical to that of RTD except for the different implant patterns. Diode-type temperature detectors were realized by selectively doping the N- and P-regions of the nanowires, using lithography. Similarly, the two contacts were doped to form N^+ and P^+ contact areas.

Measurement of the reverse-bias diode current was an excellent way to sense temperature as it varied more than an order of magnitude (at −0.5 V) for a temperature change from 293 to 373 K.

5.8 Ratiometric Fluorescent Nanoparticles for Temperature Sensing

What are the advantages of fluorescence-based temperature sensors? In the realm of temperature sensing and imaging, fluorescence-based measurements

have carved their niche because of the merits of noninvasivity, accuracy, and robustness in even strong electromagnetic fields. Fluorescence is a process distinct from incandescence (i.e., emission of light due to high temperatures). Heating is generally detrimental to the process of fluorescence, and most substances, when fluorescing, produce very little heat. Due to this reason, fluorescence has commonly been referred to as "cold light." In the standard conceptual model of molecules, the electrons occupy distinct orbits and thus energy levels. When light falls on a molecule that absorbs rather than transmits, one or more of the electrons of the molecule are "kicked" into a higher energy state. All these excited electronic states are unstable, and sooner or later the electrons lose their excess energy and fall back to lower energy states. This extra energy is dissipated in a multiplicity of ways, the most common being simply to increase atomic vibrations within the molecule. But some molecules are capable of emitting part of their energy as light. This is what we see as fluorescence.

Why are Europium (III) chelates preferred in this application? Europium (III) chelates display highly temperature-dependent emission, large Stokes' shift (the difference between excitation wavelength and emission wavelength), and long lifetime (the average time the molecule stays in its excited state before emitting a photon), which make them a strong candidate for fluorescent temperature sensing. A chelate is a chemical compound composed of a metal ion and a chelating agent. A chelating agent is a substance whose molecules can form several bonds to a single metal ion. In other words, a chelating agent is a multidentate ligand, i.e., a ligand capable of donating two or more pairs of electrons in a complexation reaction to form coordinate bonds. The ligand is a molecule, ion, or atom bonded to the central metal atom of a coordination compound.

Peng et al. (2010) reported fluorescent nanoparticles (NPs) prepared from a visible-light sensitized Eu^{3+} chelate for temperature sensing and imaging over the physiological range (25°C–45°C). The NPs displayed strong temperature dependence in both fluorescence intensity (FI) and lifetime.

What is the conceptual basis for measuring FI? The probability that a fluorophore will emit a photon is a product of the probability that a photon is absorbed and the probability that the excited state decays by emitting a photon (related to quantum yield). The probability of absorption is expressed by the molecular extinction coefficient. Molecular extinction coefficients of fluorophore are somewhat sensitive to the microenvironment, which tends to influence the energy of the electronic states. To measure FI, given a source of photons with known energies (spectral characteristics), the excitation optics will select a set of wavelengths that are incident on the sample (Gaigalas et al. 2001). The excitation optics will illuminate a specific volume inside the sample. The collection optics will collect light emitted from a volume element that we call the detection volume. The total FI after the collection optics will depend on the overlap of the illumination and detection volumes and we call these overlapping volumes the sensing volume. Only fluorophore molecules in the sensing

volume can contribute to the fluorescence signal. The FI depends directly on the concentration of fluorophore in the sensing volume, molecular extinction coefficient, and the quantum yield. The quantum yield is defined as the ratio of the total emitted fluorescent photons to the total number of absorbed photons. The collection optics will direct light of selected spectral range to the detector and provide a measurement of emission intensity over the selected range of wavelengths. Assuming that the detector is a photomultiplier tube (PMT), the PMT current is a sum of fluorescence photons of all transmitted wavelengths. The current from the PMT will be determined by PMT gain, the quantum efficiency of the PMT photocathode and the flux of fluorescence photons per unit wavelength. The PMT current is a sum of fluorescence photons of all transmitted wavelengths. Combination of the fluorescence flux per unit wavelength from the sensing volume with the current from the PMT, gives the explicit relation between the concentration of fluorophores and the measured current at the PMT detector. The relation between the fluorescence signal, in the form of current, and the fluorophore concentration, contains instrumental factors and molecular factors. It is the basis for interpreting the fluorescence signal in terms of a concentration of fluorophores.

How does a PMT work? PMTs convert photons into an electrical signal. They have a high internal gain and are sensitive detectors for low-intensity applications. A PMT consists of a photocathode and a series of dynodes in an evacuated glass enclosure. When a photon of sufficient energy strikes the photocathode, it ejects a photoelectron due to the photoelectric effect. The photocathode material is usually a mixture of alkali metals, which make the PMT sensitive to photons throughout the visible region of the electromagnetic spectrum. The current produced by incident light is multiplied by as much as 10^8 times (i.e., 160 dB), in multiple dynode stages, enabling detection of individual photons when the incident flux of light is very low.

How is fluorescence lifetime measured? There are two complementary techniques of lifetime measurement: the time domain and the frequency domain. (i) In the time domain technique, a short pulse of light excites the sample, and the subsequent fluorescence emission is recorded as a function of time. This usually takes place on the nanosecond timescale. (ii) In the frequency domain technique, the sample is excited by a modulated source of light. The fluorescence emitted by the sample has an analogous waveform. However, it is modulated and phase-shifted from the excitation curve. Both modulation (M) and phase-shift (φ) are determined by the lifetime of the sample emission. Lifetime is calculated from the observed modulation and phase-shift.

How do sensitivity and lifetime methods compare? Sensitivity of single intensity-based measurements is undermined by distribution of probes and drifts of the optoelectronic system (lamps and detectors). Although lifetime-based sensing schemes are free of these drawbacks, the complexity and the demands on the components increase with the decreasing decay times. Comparatively, fluorescence ratiometric (two-wavelength) methods are more robust and convenient in practical applications because of the built-in

calibrations provided by the simultaneous detection of two signals under one single-wavelength excitation.

How are temperature-insensitive reference dye and temperature-sensitive probing NPs used together to make a ratiometric temperature sensor? Peng et al. (2010) introduced a temperature-insensitive dye, an alkoxysilane (a reactant used for the preparation of chemically bonded phases; a compound in which silicon is attached by an oxygen bridge to an alkyl group, i.e., SiOR)-modified naphthalimide (1,8-naphthalimide = $C_{12}H_7NO_2$) derivative, N-octyl-4-(3-aminopropyltrimethoxysilane)-1,8-naphthalimide (OASN), Figure 5.7a, into the Eu^{3+} chelate-based NPs as the reference dye to the probe Eu^{3+} chelate. The temperature indicator was Eu-tris (dinaphthoylmethane) bis (trioctylphosphine oxide (Eu-DT), Figure 5.7b; trioctylphosphine oxide (TOPO) = $[CH_3(CH_2)_7]_3PO$. The strong temperature dependence of Eu-DT is attributed to the thermal deactivation of $^5D_1$ and $^5D_0$ europium energy levels through the coupling to the environmental vibration energy levels. Absorption spectra were recorded on a UV–Vis spectrophotometer. The absorption and emission spectra of Eu-DT and OASN in acetone (CH_3COCH_3) solution showed that the absorption band of the two dyes partly overlapped, which ensured ratiometric fluorescence under one single-wavelength excitation, i.e., at 400 nm. Following photoexcitation, OASN gave
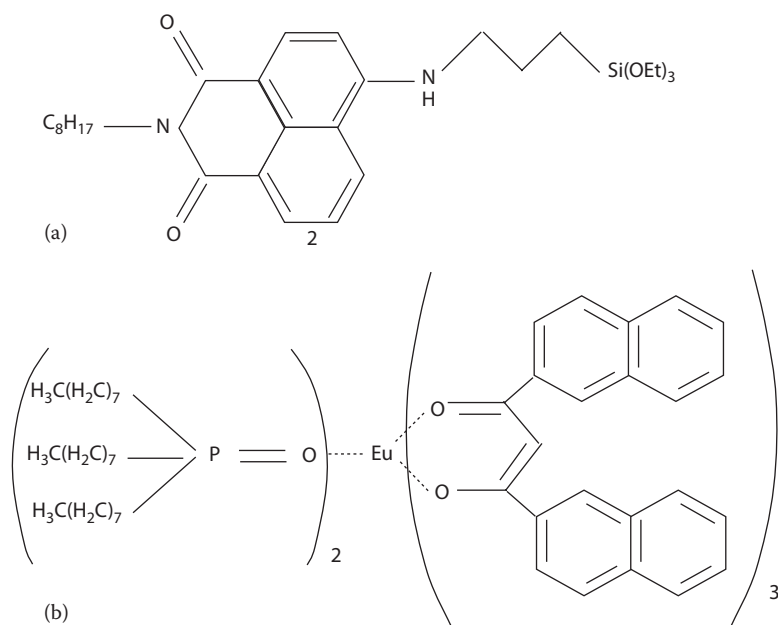
FIGURE 5.7
(a) OASN (N-octyl-4-(3-aminopropyltrimethoxysilane)-1,8-naphthalimide 2) and (b) Eu-DT (Eu-tris(dinaphthoylmethane)-bis(trioctylphosphine oxide). (After Peng, H.-S. et al., *J. Nanopart. Res.*, doi: 10.1007/s11051-010-0046-8, 2010.)
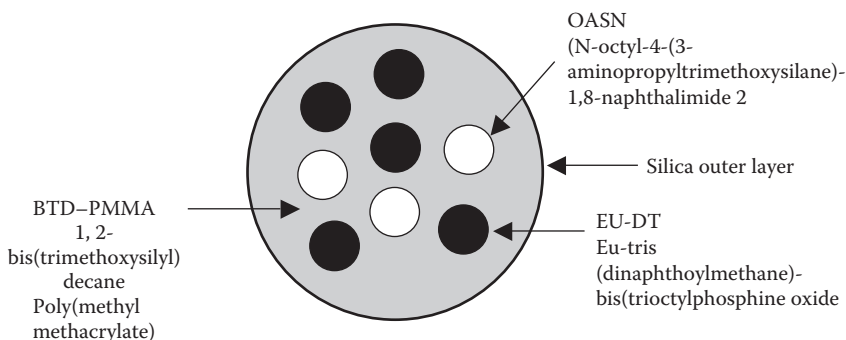
FIGURE 5.8

Ratiometric fluorescent nanoparticle. (After Peng, H.-S. et al., *J. Nanopart. Res.*, doi: 10.1007/s11051-010-0046-8, 2010.)

a green fluorescence (with a peak at 497 nm), while Eu-DT gave the characteristic red emission.

How are the ratiometric NPs made? The ratiometric NPs were prepared by the encapsulation–reprecipitation method. Eu-DT, OASN, PMMA (polymethyl methacrylate): $[C_6H_{10}O_n]_n$, and BTD (5-bromothienyldeoxyuridine: $C_{13}H_{13}BrN_2O_5S$), respectively, were used as the probe, reference dye, matrix, and silica-based encapsulation agent (Figure 5.8). The resulting NPs gave a two-wavelength emission under one single-wavelength excitation, the ratio of which is highly temperature dependent. The ratiometric NPs exhibited an intensity temperature sensitivity of -4.0% $°C^{-1}$ from 25°C to 45°C.

What are the possible applications of these nanosensors? Given the small size (20–30 nm in diameter) and biocompatible nature (silica, SiO_2 outer layer), such kind of ratiometric NPs are ideal nanosensors for sensing and imaging of cellular temperatures.

5.9 Er^{3+}/Yb^{3+} Co-Doped Gd_2O_3 Nano-Phosphor as Temperature Nanosensor Using Fluorescence Intensity Ratio Technique

Nanostructured materials for photonic applications have grabbed attention. A class of such materials is the rare earth–doped nanocrystals (REDONs). *What are rare earth elements?* Rare earth elements or rare earth metals are a group of 17 elements in the Periodic Table including scandium, yttrium, and 15 lanthanoids with Z, ranging continuously from 57 to 71 (lanthanum—La, cerium—Ce, praseodymium—Pr, neodymium—Nd, prometheum—Pm, samarium—Sm, europium—Eu, gadolinium—Gd, terbium—Tb, dysprosium—Dy, homium—Ho, erbium—Er, thulium—Tm, ytterbium—Yb, and lutetium—Lu). Scandium and yttrium are considered rare earths since they tend to occur in the same ore deposits as the lanthanoids and show similar

chemical properties. The term "rare earth" originates from the minerals from which they were first isolated, which were uncommon oxide-type minerals (earths) found in gadolinite extracted from one mine in the village of Ytterby, Sweden. Nevertheless, except for the highly unstable prometheum, rare earth elements are found in relatively high concentrations in the Earth's crust with cerium being the 25th most abundant element in the Earth's crust at 68 ppm.

REDON materials have been examined for use as phosphors in amplifiers and lasers. The promising optical properties of REDONs for photonic applications also led to the study of frequency upconversion (UC). *What are the parameters on which UC efficiency depends?* The UC efficiency depends on the nanoparticle shape, the site symmetry, and the statistical distribution of active ions. Also, the process of miniaturization of materials to the nanometer scale revealed that the radiative electronic relaxation probabilities of rare earth ions doped in dielectric nanoparticles is strikingly different from their bulk counterparts.

As nanotechnology progresses, it is interesting to enquire into the potential of REDONs as nanosensors. Temperature dependency of absorption and emission properties of rare earth–doped fluorescent materials enables their use as temperature sensors. Many researchers have considered the application of rare earth–doped bulk materials for temperature measurements using different techniques. *Are there any areas where they provide the sole solutions to temperature measurement?* They are particularly useful in many areas where electrical methods become unpractical, such as in biology and optical telecommunications.

In particular, temperature sensors based on the fluorescence intensity ratio (FIR) among different emission lines is an important method. *What is the FIR-based method?* The FIR method involves measurements of the fluorescence intensities from two closely spaced electronic energy levels that are thermally coupled and assumed to be in a thermodynamical quasi-equilibrium state (the quasi-balanced state of a thermodynamic system near to thermodynamic equilibrium). A process is termed "quasi-equilibrium or quasi-static process" if its deviation from thermodynamic equilibrium is infinitesimal). No system can be in complete equilibrium as it undergoes a real process, but some processes are almost in perfect equilibrium throughout the process.

What is the available operating range of the FIR method? The FIR method has the ability to cover a wide range (–196°C to 1350°C) with reasonable measurement resolution.

Alencar et al. (2004) reported the application of REDONs for temperature measurements using the FIR method. They studied the use of Er^{3+}-doped $BaTiO_3$ nanocrystals for thermometry and presented ensemble (a unit or group of complementary parts that contribute to a single effect) measurements on frequency upconversion of erbium ions (Er^{3+}) doped in $BaTiO_3$ nanocrystals to evaluate this material as a candidate for use as FIR-based temperature nanosensors. The fluorescence properties of the (Er^{3+})-doped $BaTiO_3$ nanocrystals were analyzed by changing the temperature of samples

made of nanocrystallites of different sizes and Er^{3+} concentrations. *How are the measurements done?* For the fluorescence measurements, the samples were placed on a hotplate and the temperature was monitored with TCs located close to the samples. Frequency-upconverted emissions centered at 526 and 547 nm from two thermodynamically coupled excited states of Er^{3+} doped in $BaTiO_3$ nanocrystals were recorded (Figure 5.9). The samples were excited with a low-power, continuous wave diode laser (wavelength: 980 nm; intensity: 2.0×10^3 W cm^{-2}) and the upconverted fluorescence, easily visible to the naked eye, was collected in the temperature range from 322 to 466 K, with a multimode fiber [a type of fiber optic cable that is thick enough for light to follow several paths through the code, mostly used for communication over short distances, such as those used in local area networks (LANs)] connected to a monochromator (an optical device that separates and transmits a narrow portion of the optical signal chosen from a wider range of wavelengths) attached to a photomultiplier. The small size of the $BaTiO_3$ nanocrystal samples (typical diameter of ~3 mm; thickness ~0.2 mm) favored their temperature homogeneity. The most common type of laser diode is formed from a P–N junction and powered by injected electric current.
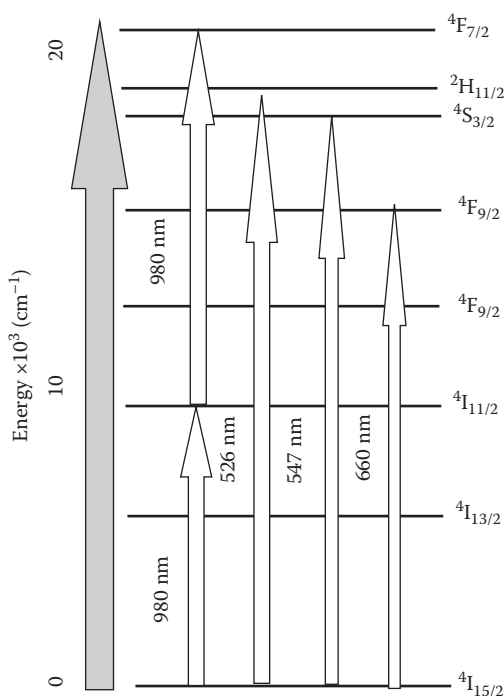
FIGURE 5.9
Simplified energy-level diagram of Er^{3+} showing the frequency upconversion scheme. (After Alencar, M. A. R. C. et al., *Appl. Phys.* Lett., 84, 4753, 2004.)

The UC signal was transmitted to a personal computer for processing. The behavior of the FIR for the emission lines centered at 526 and 547 nm was studied as a function of the temperature. The monolog plot of the experimental data yielded linear dependence of the FIR with inverse of the temperature. The data were fitted with a linear curve

$$\ln(FIR) = -\left(\frac{\alpha}{T}\right) + \beta \tag{5.5}$$

The slope of the linear curve (parameter α) is related to the sensitivity of the sensor defined as the rate in which the FIR changes with the temperature. The sensitivity found for the samples was $\leq 0.0052\,K^{-1}$. Note that α does not change, within experimental error, for the samples with different Er^{3+} concentrations and the same size but it varies when the nanoparticle size is reduced. This was explained by taking into consideration the modifications of nonradiative relaxation mechanisms (in which light is not emitted) with the size of the nanocrystals.

A temperature nanosensor based on the FIR from a single REDON will be a permanent device, illuminated with a low-cost diode laser, and will not require specific environmental conditions. The fluorescence detection scheme allows real-time reading of the temperature. One possible use of this nanosensor is for biological applications. In this case, it should be noted that the sensitivity changes when the nanocrystal is immersed in an aqueous medium because the luminescence lifetime of a REDON also changes with the surrounding medium.

Singh et al. (2009) prepared Gd_2O_3: Er^{3+}/Yb^{3+} nanocrystalline phosphor material through optimized combustion and studied upconversion excited with the 976 nm wavelength. Frequency upconverted emissions from two thermally excited states $^2H_{11/2}$ and $^4S_{3/2}$ of Er^{3+} centered at 523 and 548 nm in the phosphor were recorded in the temperature range 300–900 K. They evaluated this material as a prospective candidate for use as temperature nanosensors based on FIR.

5.10 Optical Heating of Yb³⁺–Er³⁺ Co-Doped Fluoride Nanoparticles and Distant Temperature Sensing through Luminescence

Tikhomirov et al. (2009) found that a pumping at 975 nm by laser diode has two effects: (i) a strong upconversion luminescence signal, and (ii) substantial heating of the Yb³⁺–Er³⁺ co-doped nanoparticles up to several hundred degrees. The heating and upconversion luminescence are essentially coexisting and simultaneous effects in these nanoparticles pointing out that both

effects can be used for heating of their surrounding nanovolumes and for detection of temperature rise and location of these nanovolumes.

How is the rise in temperature of nanoparticles measured? The temperature rise in the nanoparticles is estimated on the basis of measurements of the intensity ratio of upconversion luminescence bands $^2H_{11/2} \rightarrow {}^4I_{15/2}$ and $^4S_{3/2} \rightarrow {}^4I_{15/2}$ of Er^{3+}. The upconversion visible luminescence allowed distant optical detection of heated nanoparticles and their surrounding nanovolumes.

What is the measurement procedure? The Yb^{3+}–Er^{3+} co-doped nanopowder was kept either in ampoules sealed under vacuum or was free-standing in the air while fixed between silica glass plates. The luminescence spectra were excited by a laser diode operating at 975 nm and up to 260 mW power. When pumped by a 975 nm laser diode into absorption band of Yb^{3+}, a laser-induced temperature rise up to 800°C was detected in the nanoparticles by measuring the ratio of the intensities of the thermalized upconversion luminescence bands.

If I_H and I_S stand for integral intensity of the $^2H_{11/2} \rightarrow {}^4I_{15/2}$ and $^4S_{3/2} \rightarrow {}^4I_{15/2}$ of Er^{3+} emission bands, respectively

$$\ln\left(\frac{I_H}{I_S}\right) = C - \frac{B}{T} \tag{5.6}$$

where B and C are the constants, $B \approx 1100\,K$, while C varies between 1.5 and 2.5. Hence the ratio I_H/I_S is defined by the temperature T of emitting sample. However, the exact values of constants B and C should be experimentally found for a specific sample before Equation 5.6 could be applied to evaluate the sample temperature precisely. Therefore, first they measured a temperature dependence of the ratio I_H/I_S versus sample temperature T when these emission bands were pumped via the higher laying $^4F_{3/2}$ absorption level of Er^{3+}.

Where are such heaters and sensors useful? The aforementioned nanoheater and temperature nanosensor may be used in medicine for local hypothermal treatment of cells and for perforation of nanoholes in organics and metals.

Example 5.2

Find the integral intensity ratio I_H/I_S at $T = 700$ and $900\,K$ for Yb^{3+}–Er^{3+} co-doped nanoparticles, assuming $B \approx 1100\,K$ and $C = 2$.

At $700\,K$, using Equation 5.6

$$\frac{I_H}{I_S} = \exp\left(C - \frac{B}{T}\right) = \exp\left(2 - \frac{1100}{700}\right) = \exp(0.4286) = 1.535 \tag{5.7}$$

Similarly at $900\,K$, the I_H/I_S ratio is obtained as

$$\frac{I_H}{I_S} = \exp(2 - 1.22) = 2.18 \tag{5.8}$$

5.11 Porphyrin-Containing Copolymer as Thermochromic Nanosensor

What is the present universal approach for obtaining desirable multicolors? It is the use of nanocrystals such as gold (Au), silver (Ag), and semiconductor quantum dots. However, tunability of the color of these nanocrystals offers a limited flexibility. Moreover, they are generally synthesized under harsh conditions and suffer from cytotoxicity and leakage into the biological system. *How is this problem solved?* The problem can be solved by using stimuli-responsive chromatic (relating to colors or color) materials as an alternative route to fulfill the color modulation. Although great efforts were devoted to the simple- or multiple-responsive chromatic switches, the study of soluble nanosensors that have distinct color transitions and broad color-tuning range (>three kinds) is still a hot research area.

The porphyrin is any of a class of water-soluble, nitrogenous biological pigments (biochromes), derivatives of which include the hemoproteins (porphyrins combined with metals and protein). A porphyrin is an organic compound that contains four pyrrole rings. A pyrrole is a pentagon-shaped ring of four carbon atoms with a nitrogen atom at one corner (C_4H_5N). The porphyrins are actually an enormous group of organic compounds, found all over the living world. They are universal, found in most living cells of animals and plants, where they perform a wide variety of functions. The special property of porphyrins is that they bind metals. The four nitrogen atoms in the middle of the porphyrin molecule act as teeth: they can grab and hold metal ions such as magnesium (Mg), iron (Fe), zinc (Zn), nickel (Ni), cobalt (Co), copper (Cu), and silver (Ag).

Based on the principle of specific chromophore (a chemical group capable of selective light absorption resulting in the coloration of certain organic compounds) location and polymer phase induction, Yan et al. (2010) reported a full-color polymer optical nanosensor, possessing four prominent advantages: (i) Metal-ion triggering of a full-spectrum-tunable sensing pattern that can enable its use as an ion detector and as a light colorimeter (device for measurement of color) at ambient temperature. (ii) Each ion-containing polymer sensor possesses a distinct thermochromic temperature, hence enabling a broad range as an ultrasensitive thermometer; "thermochromism" is the ability of substance to change color due to a change in temperature. (iii) Dispersive polymeric nanoparticles with high water solubility acting as sensing units, maintaining a discrete signal transport. (iv) Full-color modulation so that the sensor is capable of sensing multichannel signals. (v) Low cytotoxicity for acceptability in biosystems.

They designed and synthesized a specific ABC triblock copolymer. Block copolymers are material systems in which a chain of N_A units of monomer A is covalently linked to a chain of N_B units of monomer B. Difunctional initiators enable polymerization by addition at both the ends of B block leading

to triblocks such as ABA, ABC, etc. *What interesting properties does this polymer display?* Interest in triblock copolymers has increased due to their complexity and physical properties. Below the low critical soluble temperature (LCST) of poly(*N*-isopropylacrylamide) (PNIPA) hydrogel (~32°C), a transparent red–brown solution with a diagnostic tetra(4-carboxylatophenyl) porphyrins (TCPP) Soret band ($\lambda_{max} = 434$ nm) was observed; meso-tetra(4-carboxyphenyl)porphine ($C_{48}H_{30}N_4O_8$) Soret band is a very strong absorption band in the blue region of the optical absorption spectrum of a heme protein. Upon heating above 35°C, interestingly, a clear blueshift from 434 to 406 nm was seen and was accompanied by a visible turbid transition (transmission: 100%–20%) by UV–Vis monitoring, implying an aggregated state transition around TCPP species.

To ensure the thermochromism of the copolymer, they investigated as to how to further enlarge the stimuli-responsive full-color range. Below 32°C, an addition of different metal ions to the solution leads to an unprecedented full spectral color range with the Soret $\pi \rightarrow \pi^*$ absorption of the TCPP species showing a large variation from 418 to 512 nm upon metal ion coordination with TCPPs. Even though the metalloporphyrin (a combination of a metal with porphyrin, as in heme) had a different color than the porphyrin, a full-spectrum color-tuning sensor by cation-stimulus TCPP-containing block copolymer was shown for the first time. Most surprisingly, on heating these multicolor nanosensors, they displayed discrete thermochromic characteristics in the temperature range of 35°C–61°C with the phase transition point (the value of the temperature, pressure, or other physical quantity at which a phase transition, change of a substance from one phase to another [solid–liquid–gas], also called a phase transformation, occurs is called a transition point) dependent on the metal ion.

The synthesis and evaluation of responsive copolymers bearing TCPP species demonstrated that it possesses conceivable thermochromism in the absence of metal ion stimuli. Upon modification by metal ions, the available color range was hugely extended and upon heating the solutions, color transitions at different thermochromic points were observed. It is anticipated that this polymeric optical nanosensor can give new insights into full-spectrum colorimetric and ultrasensitive thermometric arrays.

5.12 Silicon-Micromachined Scanning Thermal Profiler

The scanning tunneling microscope (STM) and the atomic force microscope (AFM) have found widespread use, and are considered important milestones in the development of scanning probe microscopes (SPMs). The spatial resolution achievable by either one of these techniques is 1 Å, while the operating gap between the tip and the sample is typically 10 Å.

What is the principle of STP? Similar in principle to the aforementioned instruments, the scanning thermal profiler (STP) interacts with the sample by sensing heat conducted through the scanning probe. It provides both temperature and topographic information (graphical representation of the surface features). Although the best reported spatial resolution for it is about an order of magnitude larger than for AFMs and STMs, it permits wide latitude in the gap size. *Why?* This is because the thermal interaction between the tip and the sample exists over a much longer range than tunneling current and near-field forces.

What is the construction of STP and how does it function? The STP is basically a miniature TC, which is thermally biased with respect to the sample and separated from it by a small air gap (Figure 5.10) (Gianchandani and Najafi 1997). In a closed-loop operation, the voltage of the TC is stabilized by a feedback system that adjusts the gap such that the tip traces a contour of constant temperature (CT). If the variation in temperature across the sample surface is small compared to the thermal bias, the contour provides a topographic map of the sample.

What are the operational modes of STP? There are two scan modes in which the STP is operated: the DC mode and the dithered (a state of indecisive agitation) mode. In the former, the tip is rastered across the sample while the air gap is controlled to keep the TC signal constant. In the latter, the tip
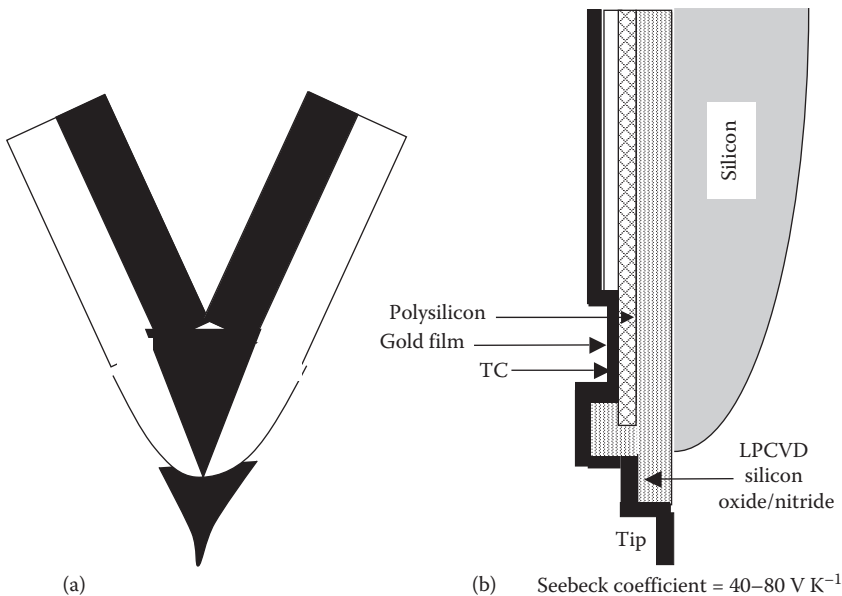
FIGURE 5.10
Surface profiler tips: (a) tungsten–nickel thermocouple <100 nm diameter and (b) polysilicon–gold thermocouple junction tip. (After Gianchandani, Y. B. and Naja, K., *IEEE Trans. Electron Devices*, 44, 1857, 1997.)

is oscillated perpendicular to the sample, modulating the air gap by about 10%. *What is the advantage of an oscillating tip?* Dithering the tip reduces sensitivity of the TC signal to fluctuations in the ambient temperature and also improves the signal-to-noise ratio, a ratio of desired signal to undesired signal (noise), serving to quantify how much a signal has been corrupted by noise.

The scanning probe overhangs the glass substrate, and is integrated with a suspension, which includes flexible support beams and electrostatic comb drives for longitudinal actuation. A thin film polysilicon–Au TC runs along the probe, with one junction near the protruding metal tip and another near the suspension. Polysilicon–gold TCs have a Seebeck coefficient (the ratio of the open-circuit voltage to the temperature difference between the hot and cold junctions) of 40–80 μV K^{-1}. A polysilicon resistive heater located on the lower surface of the suspension is used for generating the thermal bias between the tip and the sample. The approach selected for fabricating the SMTP is the bulk silicon micromachining technique.

What parameters determine the spatial resolution of the tip? Spatial resolution of an SMTP depends on the Seebeck coefficient of the TC, the thermal impedance of the probe shank, as well as on operating conditions such as temperature bias, air gap, and the sensitivity of the interface electronics. For a $1\,\mu m \times 0.5\,\mu m$ tip and a $0.1\,\mu m$ long air gap, the spatial resolution and the device noise equivalent temperature difference (NETD) have been theoretically estimated as ~3.33 nm and ~0.1 mK $Hz^{-1/2}$, respectively. NETD is a measure of the sensitivity of a detector of thermal radiation. It specifies the amount of radiation required to produce an output signal equal to the detector's own noise (due to inner component heat). Thus, it specifies the minimum detectable temperature difference. In general, detector cooling is required to limit the detector's own noise and to improve the NETD.

5.13 Superconducting Hot Electron Nanobolometers

What is a bolometer? A bolometer is a device for measuring the energy of incident electromagnetic radiation. It measures electromagnetic radiation in its various forms, from radio waves (30 kHz–300 GHz) to ultraviolet radiation (7.5×10^{14}–3×10^{16} Hz) and gamma rays (typically $>10^{20}$ Hz). Invented by American astronomer Samuel Pierpont Langley in 1878, the first bolometer was used in conjunction with a telescope to measure infrared radiation (0.003–4×10^{14} Hz) on astronomical objects, namely, the Moon. Since the original invention, bolometers have treaded a long way of improving the sensitivity and expanding the frequency range, from x-rays (3×10^{16} Hz upward) and optical (4–7.5×10^{14} Hz)/UV(7.5×10^{14}–3×10^{16} Hz) radiation to the submillimeter waves (3×10^{11}–3×10^{12} Hz or 3 THz); the latter range contains 50% of

the total luminosity of the universe and 0.98 of all the photons emitted since the Big Bang occurred.

What is the construction of the bolometer? A bolometer consists of an absorptive element, such as a thin layer of metal or metallic strip, connected to a heat sink (a body of CT) through a thermal link. Hence, any radiation impinging on the absorptive element, i.e., the metallic strip, raises its temperature above that of the heat sink. The more the energy absorbed, the higher is the temperature. The temperature change is measured directly or via an attached thermometer.

Metal bolometers, produced from thin foils or metal films, usually work without cooling. Today, most bolometers use semiconductor or superconductor (materials that have zero electrical resistivity, and the magnetic field inside a bulk sample is zero [the Meissner effect]) absorptive elements instead of metals (Romestain et al. 2004). Superconducting photodetectors offer an interesting alternative to traditional photon counting systems. In these superconducting devices, different mechanisms are exploited to detect the photon absorption, depending on the type of detector (transition edge sensor [TES]), superconducting tunnel junction, and hot-electron bolometer (HEB).

What is a hot-electron bolometer? An HEB is a detector for submillimeter radiation and far infrared (Cherednichenko and Drakinskiy 2008, Gerecht and You 2008, Jiang et al. 2009). The detector is mainly interesting for detecting radiation with a frequency of above 1000 GHz, an area where heterodyne (having alternating currents of two different frequencies that are combined to produce two new frequencies, the sum and difference of the original frequencies, either of which may be used in receivers by proper tuning or filtering; applied to frequency shifting information of interest into a useful frequency range) SIS (superconducting–insulating–superconducting) receivers no longer work well. A miniscule but supersensitive sensor can help solve the mysteries of outer space. Cosmic radiation (energetic charged subatomic particles, originating from outer space, which impinge on the Earth's atmosphere), which contains the terahertz frequencies (3×10^{11}–3×10^{12} Hz) that the sensors detect, offers astronomers important information about the birth of star systems and planets.

Like the heterodyne SIS receiver, the HEB makes use of heterodyne (Greek roots hetero- "different," and dyn- "power") mixing to transform the received signal to a lower frequency. The mixed signal has a frequency of a few gigahertz. The bolometer can follow this mixed signal exactly. The frequency, phase, and power of the radiation absorbed can be deduced from the electronic signal.

What are nanobolometers? These are based on thin film metal nanostructures that work at ultralow temperatures (Vystavkin et al. 2010). HEB detectors made of thin superconducting films (≈5 nm thick) of niobium nitride (NbN) have been found to present the best figure of merit for detecting single visible optical photons (Figure 5.11). The core of the detector comprises a small piece
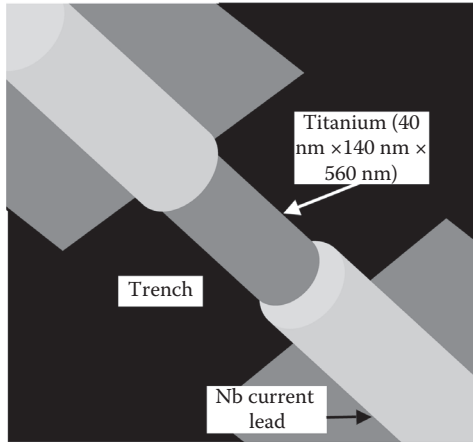
FIGURE 5.11
Ti nanosensor. (After Vystavkin, A. N. et al., *J. Commun. Technol. Electron.*, 55, 710, 2010.)

of superconducting NbN; detectors based on it can detect a single photon in the 3–10 μm section of the infrared spectrum. Clean superconducting contacts that are kept at a CT of –268°C (5° above absolute zero) are attached to both ends of the superconducting NbN. Miniscule gold antenna catches the terahertz-radiation and sends it via the contacts to the small piece of NbN, which functions as an extremely sensitive thermometer. Thus the bolometer consists of the superconducting NbN nanobridge at the center, which connects to the on-chip gold spiral antenna via additional contact pads.

What does the term "hot electron" mean in reference to the bolometer? Due to the low temperature of the strip, the incoming radiation first of all heats up the electrons in the NbN nanobridge. The electron gas in the volume of the nanobolometer serves as the absorber. As the electrons of the sensor absorb radiation, the temperature of the electrons rises above that of the sensor itself. In superconducting thin films, the electrons are only loosely coupled to the crystal lattice and the electron–electron interaction is enhanced. When bombarded by photons, these electrons equilibrate at a temperature greater than the lattice. The electrons then cool either through phonon coupling to the underlying lattice or through diffusion across normal metal contacts. Principally, it is the length of the bridge that decides which cooling mechanism dominates. Diffusion cooling (the situation in which the average energy of an electron swarm is reduced by the diffusion of the faster electrons to the container walls) becomes prominent as the bridge length becomes smaller.

Due to the heating of free electrons by the absorption of photons, their temperature locally becomes higher than the superconducting transition temperature (the critical temperature at which the electrical resistivity of a metal drops to zero) of the NbN material ($T_C = 15.7$ K). These devices are therefore

referred to as hot electron bolometers (HEBs); the bolometer derives its name from this characteristic.

What is a transition edge sensor? Superconducting TES are biased close to the superconducting-normal metal transition edge to detect the large increase in resistance as the photon energy is absorbed. When biased on the superconducting transition, the device has a finite electrical resistance that is less than the resistance in the fully non-superconducting state. Energy coupled to the detector increases the temperature of the superconducting material, pushing it further into the non-superconducting state and thereby increasing its electrical resistance. The nanobridge suffers from a localized loss of its superconducting characteristics and it detects the radiation used to detect very small changes in temperature, and hence in energy.

The temperature of the electrons in the strip must be able to chase the wave movement of the mixed signal exactly, more than 10^9 times per second; the electrons are therefore slightly warmed. For continued measurements, the electrons in the strip must be cooled down to the superconducting temperature extremely fast. *Is it possible to cool the electrons?* Yes, because the electrons rapidly release their heat to the atoms of the NbN film. As the speed of this process is directly dependent on the film thickness, the NbN is extremely thin; 5 nm is just a few hundred atomic layers.

Ultrasensitive arrays based on superconducting nanobolometer sensors are used for the passive radars (radars that have one or more receivers but no active transmitter, i.e., radars that receive only, instead of alternating transmission and reception) at the terahertz frequencies (0.3–10 THz) including the imaging radiometers for the ground-based and space astronomical complexes, security systems (detection of hidden weapons, explosives, etc.), medical diagnostics, environmental control, diagnostics of various products (including food products), etc. *What applications require the highest sensitivity?* The highest sensitivity is needed for the space telescopes (instruments launched into orbit in outer space which are used for observation of distant planets, galaxies, etc.), in which the fundamental limit is related to the quantum noise (uncertainty of some physical quantity due to its quantum origin, attributable to the discrete and probabilistic nature of physical phenomena and their interactions, and representing the fundamental limit of the achievable signal-to-noise ratio of an optical communication system) of the background space signal and the received radiation. The noise per bolometer is $\sim 10^{-21}$ W/Hz$^{1/2}$ in the terahertz frequency range. The nanobolometer with a working temperature of about 40 mK and dimensions (length, width, and thickness) of 1.0, 0.13, and 0.04 µm is predicted to have a sensitivity close to the limiting sensitivity.

Wei et al. (2008) fabricated the nanobolometers using electron-beam lithography and electron-gun deposition of Ti and Nb films on Si substrates covered with a film of SiO_2 or Si_3O_4. The device consisted of a titanium "island" with a volume of $\sim 10^{-2}$ µm^3 flanked by niobium current leads. Low resistance of the Ti/Nb interface was ensured through *in situ* deposition of Ti and Nb

at different angles through a "shadow mask." The critical temperature T_C (Ti) = 0.1–0.2 K for thin ($d \approx 40$ nm) and narrow ($w \approx 0.1$ μm) Ti nanobridges was lower than the bulk value (0.4 K). These devices operated at sub-Kelvin temperatures in the hot-electron regime at the temperature $T < T_C$ (Ti). Absorption of radiation by electrons in the titanium island increased their effective temperature T_e and caused a corresponding change in resistance while the crystal lattice remained in equilibrium with the bath. Superconducting leads with high critical temperature severely increased thermal isolation of electrons in the nanosensor.

5.14 Thermal Convective Accelerometer Using CNT Sensing Element

This device works as an anemometer, a device for measuring the speed of airflow in the atmosphere, in wind tunnels, and in other gas-flow applications through temperature sensing. Structurally, the CNT-based thermal convective motion sensor (Zhang et al. 2009) consists of a sensor chip and a chamber to seal the convection medium. The sensor chip was fabricated using conventional micro fabrication method along with dielectrophoresis (DEP: the translational motion of uncharged suspended particles relative to that of the suspending medium when the suspension is subjected to a nonuniform electric field) manipulation of CNTs. A 1 mm thick glass was selected as the substrate. The forced convection (the transference of heat by the flow of fluids such as air or water, driven by fans, blowers, or pumps) in the chamber generated by vibration was sensed by the CNT bundle.

The heat-generating CNT bundle was first heated up in a micro chamber using constant current. External acceleration caused thermal convection (heat transfer by the combined mechanisms of fluid mixing and thermal conduction) within the chamber, which was detected by the fluctuation of the output voltage of the CNT temperature-sensing element.

Two temperature sensors were placed symmetrically around the heater along the sensitive axis. Under acceleration, the temperature distribution around the heater became asymmetric with the heater due to induced convection. The acceleration was derived since it was proportional to the measured temperature difference between the two sensors.

What are the advantages that favor this method? By using this detection method, the sensing block consumes only tens of pico-watts (10^{-12} W) to function as motion sensors. Further, by using this sensing method, only micro heater and temperature detectors are needed and no moving parts (proof-mass) are used. So this type of accelerometer is more robust to shocks compared to conventional micro motion sensors.

5.15 SWCNT Sensor for Airflow Measurement

The heat transfer between the CNT and the environment is influenced by the rate of fluid flow around the CNT. This heat transfer in turn affects the power required to heat the CNT to a specified temperature; hence, a CNT carefully biased for self-heating is usable as a fluid flow rate sensor.

What are the special advantages offered by CNT-based sensors for fluid flow studies? CNT-based fluid flow sensors are easily self-heated so that they have microwatt power consumption compared with milliwatt power consumption of polysilicon-based resistors. Moreover, CNT fluid flow sensor can be used at a scale of hundreds of micrometers within a micro wind tunnel, a chamber through which air is forced at controlled velocities and in which airplanes, motor vehicles, etc., or their scale models, are tested to determine the effects of wind pressure on aerodynamic flow around airfoils, or other objects. Fluid flow rate inside the tunnel is increased due to compression and the responsivity of the sensor is thus increased.

Chow et al. (2010) developed single-walled carbon nanotube (SWCNT) sensors for airflow shear-stress (the form of stress in a body, which is applied parallel or tangential to a face of a material, as opposed to a normal stress that is applied perpendicularly, that causes parallel layers of a material to move relative to each other in their own planes and tends to produce cutting rather than stretching or bending) measurement inside a PMMA (($C_5O_2H_8)_n$, a transparent, thermoplastic polymer derived from methylacrylate and used as a substitute for glass) micro-wind tunnel chip. This shear-stress sensor consisted of a thermal element. The element was located on the surface of the substrate and possessed a pronounced temperature dependence of resistivity. The thermal element resided within a velocity boundary layer. In this layer, the velocity changed from zero, at the wall, to the value of the mean stress flow. When current was applied to the thermal element, Joule heating ($Q = I^2Rt$, where Q is the heat generated by a constant current I flowing through a conductor of electrical resistance R, for a time t) took place raising the temperature of the thermal element. But when the air flow was introduced on to the heated element, the temperature of the heated element decreased. This was caused by the interaction between the flow and the heated element. The resistance decreased if the element had positive TCR or increased if it had a negative TCR. The rate of heat loss from a heated resistive element to the ambient was dependent on the velocity profile (the shape of the velocity curve) in the boundary layer. Since the sensor was operated in CT mode, when the sensor lost heat to the surrounding flow, its resistance changed accompanied by a current flow to the sensor to keep the sensor resistance constant with input power of ~230 μW. The voltage output of the sensor increased with the increasing flow rate in the micro-wind tunnel and the measurable volumetric flow (the volume of the fluid that passes through a given surface per unit time) was in the order of 1×10^{-5} m^3 s^{-1}.

An array of sensors was fabricated by employing dielectrophoretic technique to manipulate bundled SWCNTs across the gold microelectrodes on a PMMA substrate. After the nanotubes were manipulated across the Au microelectrodes by DEP, the sensor was annealed by a high current for several cycles. The intent was to burn off those SWCNTs that had a weak adhesion with the Au microelectrodes. The sensors were then integrated in a PMMA micro-wind tunnel, which was fabricated by SU-8 molding/hot-embossing technique. The SWCNTs sensors showed a negative TCR of approximately -0.25% °C^{-1}.

5.16 Vacuum Pressure and Flow Velocity Sensors Using Batch-Processed CNT Wall

The pressure and flow sensing (Choi and Kim 2010) relies on three main mechanisms: (i) Negative TCR of the CNTs. (ii) The thermistor effect of an electrothermally heated CNT wall; thermistors are a type of resistor composed of solid semiconducting materials, with an electrical resistance that possesses either a negative or positive temperature coefficient of resistivity. (iii) The temperature-dependent tunneling rate at the CNT–silicon interface.

How does the sensor work? When the supply of constant electrical energy on the CNT wall is maintained, the temperature of the CNTs depends on the loss of thermal energy via heat transference through the surrounding gas molecules. Hence, the pressure and flow velocity change causes a temperature change in the CNTs and at the interface between the CNTs and silicon that is monitored by measuring the resistance change. The sensing capability of CNT wall extends across a wide pressure range.

How was the sensor fabricated? For sensor fabrication, a CNT wall was synthesized using thermal CVD on the sidewall of a microelectrode array fabricated by a simple micromachining process on an SOI wafer. The synthesized CNT wall was a bamboo-shaped structure consisting of an MWCNT, as observed through TEM analysis. The temperature dependence of the resistance of the CNT wall was measured, and the extracted TCR was -0.84% K^{-1}. The sensor responded immediately to a change in pressure and flow velocity. The maximum detectable range of pressure was between 10^5 and 3×10^{-3} Pa, and the range of flow velocity tested was from 1 to 52.4 mm s^{-1} in a nitrogen environment.

What are the benefits of the batch process and how does it differ from existing methods? The batch fabrication process presented by the workers, included synthesis, integration, and assembly of CNTs. It was quite simple and resulted in a high yield in opposition to existing methods for CNT-based sensors. Contrary to the proposed method, the traditional manufacturing methods for CNT sensors entail fairly low throughput techniques such as

dielectrophoretic placement of CNTs, dispersion of pre-synthesized CNTs onto the substrate, and electron-beam lithography patterning. Although these conventional approaches are efficient in integrating an individual CNT or a few CNTs between electrodes, the labor-intensive process is serial, time consuming, and uneconomical, thereby encumbering the mass production of CNT-based sensors.

In the batch fabrication process, the microelectrodes were fabricated on a SOI wafer by photolithography and deep reactive ion etching (DRIE) of the silicon device layer. The CNTs were synthesized using thermal CVD to form CNT walls around an array of microelectrodes. During the subsequent electron-beam evaporation, the catalyst metal (Fe) was deposited on the top of electrodes and substrate and on the sidewall of the electrodes due to imperfect directionality, even though the deposited thickness at the sidewall was thinner than on the top. By a reactive ion etching (RIE) step, the Fe catalyst was removed from the top of the electrodes and substrate.

Why and how is the process amenable to mass production? CNT-based 2T devices were realized through the CNT wall formed and self-assembled between the electrodes and substrate along the sidewall of the electrodes without additional post-assembly processing. After synthesizing of CNTs, no CNTs were observed on the topside of the electrodes and substrate. This revealed the complete etching away of the catalyst metal on the topside of the electrodes during the RIE process. This synthesis of CNTs at selected locations was responsible for the bulk production capability of the process.

Since batch-processing was involved in all the fabrication steps, including the formation of microelectrodes and the synthesis and assembly of CNTs, high throughput and yield were achieved, contrasting with fabrication of most of the existing CNT-based sensors.

5.17 Nanogap Pirani Gauge

The miniaturization limits of classical diaphragm-based pressure sensors are decided by the technological limits on diaphragm fabrication, junction depth, etc. Although these limits have continuously improved, the physical problem of sensing very small diaphragm stresses or displacements still persists. Pirani-based pressure sensors have no moving part, they are relatively easy to fabricate and can be extremely miniaturized. The traditional Pirani gauge (range: 1–10^{-4} torr, 10^2–10^{-2} Pa) is a thermal conductivity gauge (a gauge measuring 100–10^{-4} torr pressure and working on the principle that thermal energy dissipation is affected by the pressure and thermal conductivity characteristics of the residual gas) containing two filaments placed along the two arms of a Wheatstone bridge. One heated filament ($120°C$–$200°C$) is exposed to the vacuum in the chamber while the other cold filament acts as a *reference*

or compensator to take care of ambient temperature fluctuations. The temperatures of the two filaments are kept constant by adjusting the power delivered to the filaments. The power fed to the measuring filament is related to the pressure in the chamber, because it is inversely proportional to the amount of gas molecules striking the filament and absorbing thermal energy from it.

Pirani gauge with a range above atmospheric pressure is a potential substitute for diaphragm-based pressure sensor at moderate atmospheric pressures. The miniaturization of Pirani gauge by micromachining provides important advantages including a small size, low fabrication cost, low power consumption, higher measurement sensitivity, fast response time, and improved dynamic range (the ratio between the largest and smallest possible values of a changeable quantity) of operating pressures.

A micromachined Pirani-based pressure sensor consists of a heater suspended over a heat sink (substrate) with a small gap between them; the heat sink is a component designed to lower the temperature of an electronic device by dissipating heat into the surrounding air. To extend the upper pressure limit of a Pirani gauge into the atmospheric range, the gap size of the device has to be as small as possible. The smaller gap size pushes the transition from molecular to continuum heat conduction into higher pressures. The transition pressure of the gauge is an empirical value that distinguishes the continuum and molecular heat conduction regimes of a gas. The continuum gas conduction regime occurs where the gas molecules carrying the heat collide with each other in the process of traveling. When the vacuum pressure is relatively low, heat conduction is by free molecule gas conduction. Above the transition pressure, a Pirani gauge loses its sensitivity and the gauge output gradually saturates.

Puers et al. (2002) used FIB-based prototyping method for fabricating a nano Pirani gauge involving a combination of FIB deposition and milling (the process of cutting away solid material). The sensor consisted of a suspended tungsten microbridge with a submicron gap from the substrate (Figure 5.12). The tungsten bridge resistor is connected to preprocessed aluminum bonding pads. The active area of the sensor was $10 \times 1\,\mu m$. In the nano gauge fabrication process, the deposition of metallic structural layers and supporting features occurred by a local CVD reaction induced by the ion beam while milling took place by scanning a focused beam of energetic Ga^+ ions causing sputtering from the surface.

Khosraviani and Leung (2009) fabricated a surface micromachined Pirani pressure sensor with an extremely narrow gap ~50 nm between its heater and heat sink (substrate) with superior output linearity in the atmospheric pressure range. The gap size of the device has been reduced to 50 nm by using a layer of PECVD amorphous silicon as a sacrificial layer and a xenon difluoride (XeF_2) gas phase etching technique. Such a slender gap pushes the transition from molecular to continuum heat conduction to pressures beyond 200 kPa. Theoretically, when the gap size approaches the size of gas molecules, the device sensitivity decreases drastically as the gap will be
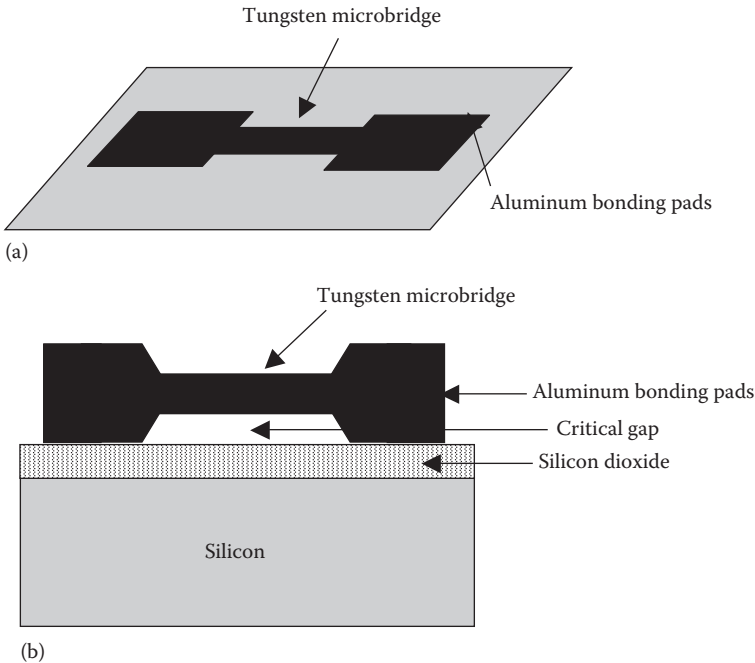
FIGURE 5.12

(a) Top view and (b) cross section of nano Pirani gauge. (After Puers, R. et al., *Sens. Actuators, A*, 97, 208, 2002.)

devoid of gas molecules responsible for heat conduction. This lower limit of the gap dimension is less than 1 nm, which is about two orders of magnitude smaller than the gap size achieved.

5.18 Carbon Nanotube–Polymer Nanocomposite as Conductivity Response Infrared Nanosensor

How is the IR photoresponse of SWCNTs affected by embedding them in an insulating matrix? Pradhan et al. (2008) reported that the IR photoresponse of SWCNTs was dramatically improved by embedding SWCNTs in an electrically and thermally insulating polymer matrix such as PC (a thermoplastic polymer containing carbonate groups (–O–(C=O)–O–)) in the air at room temperature (Figure 5.13). *How does the response of embedded SWCNTs differ from that without embedding?* HiPco-produced SWCNT ($SWCNT_{HiPco}$) film (power intensity: 7 mW mm^{-2}; on/off time period: 200 s) exhibited two prominent features: (i) weak conductivity change (1.10%) upon IR illumination and (ii) gradual rise and decay of the conductivity with the on/off IR illumination. In contrast to
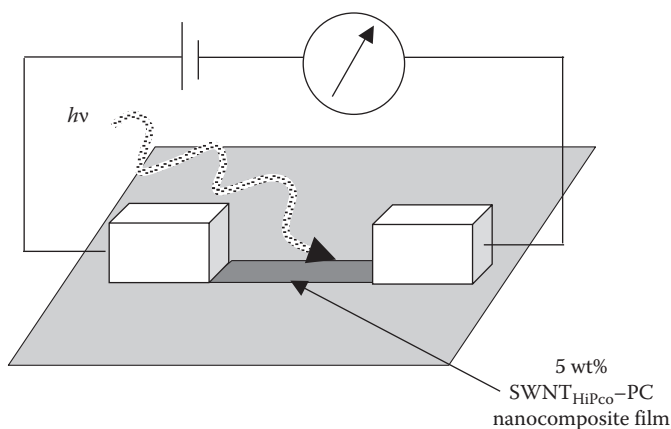
FIGURE 5.13
Set up of CNT-polymer nanosensor for conductivity measurements. (After Pradhan, B. et al., *Nano Lett.*, 8, 1142, 2008.)

the gradual photoresponse and feeble conductivity change (1.10%) observed in a $SWCNT_{HiPco}$ film in the air at room temperature, the 5 wt% $SWCNT_{HiPco}$–PC nanocomposite under the same IR illumination (power intensity: 7 mW mm^{-2}; on/off time period: 30 s) showed a sharp photoresponse and pronounced conductivity change (4.26%), which was ~4 times of that (1.10%) observed in the pure $SWCNT_{HiPco}$ film under the same condition, 21 times of that (0.2%) in the pure $SWCNT_{arc}$ film in the air at room temperature (12 mW mm^{-2}), and 9 and 6 times of that (0.7%) in the pure $SWCNT_{arc}$ film in the vacuum at 50 K. Even the 1 wt% $SWCNT_{HiPco}$–PC nanocomposite film gave a remarkable conductivity increase of 2.56% upon the same IR radiation, which surpassed that of pure $SWCNT_{HiPco}$ film (1.10% conductivity change) under the same conditions. While the IR photoresponse of the pure $SWCNT_{HiPco}$ film was predominated by thermal effect, in the IR photoresponse of the 5 wt% $SWCNT_{HiPco}$–PC nanocomposite, photoexcitation effect was preponderant. Photoexcitation is the mechanism of electron excitation by photon absorption, leading to the formation of an excited state of atoms or molecules, when the energy of the photon is too low to cause photoionization.

How are the SWCNTs embedded in PC matrix? Purified $SWCNTs_{HiPco}$ were solubilized in chloroform ($CHCl_3$) with standard personal protective equipment (PPE) by vigorously shaking and/or short bath sonication. The $SWCNT_{HiPco}$ solution thus formed was mixed with a PC solution (for undoped samples) or a PC-iodine(I) solution (for doped samples) in chloroform to produce a homogeneous $SWCNT_{HiPco}$–PC composite solution. This composite solution was cast on a glass dish and dried very slowly to give a free-standing film after peeling from the substrate. The aforementioned process was followed to prepare doped and undoped $SWCNT_{HiPco}$–PC composites with various $SWCNT_{HiPco}$ loadings. The pure $SWCNT_{HiPco}$ film was obtained from the

PTFE [polytetrafluoroethylene: $(C_2F_4)_n$] membrane after filtration of a suspension of SWCNTs$_{HiPco}$ in chloroform. The typical film thickness was in the range 25–60 µm. The mass ratios of PPE:SWCNT$_{HiPco}$ and iodine:SWCNT$_{HiPco}$ were kept at 0.4 and 5, respectively. SEM showed the excellent dispersion of PPE–SWCNTs$_{HiPco}$ in the PC matrix, which was vital for obtaining composites with isotropic electrical conductivity.

What are the likely effects of photoexcitation on the conductivity of SWCNTs? The photoexcitation of semiconducting SWCNTs with IR light leads to the generation of excitons (a fundamental quantum of electronic excitation in condensed matter, i.e., matter in the liquid or solid state, consisting of a negatively charged electron and a positively charged hole bound to each other by electrostatic attraction; a bound state of an electron and hole) instead of free carriers. The IR light could have two main effects on the conductivity of SWCNTs: (i) *Photo effect:* The excitons can be dissociated into free electrons and holes thermally or by a large electric field, which increases the conductivity of SWCNTs. (ii) *Thermal effect:* The excitons decay into heat and the strong warming of SWCNT reduces its resistance.

5.19 Nanocalorimetry

Thus far, attention has been focused on temperature sensing, but temperature is the effect and heat is the underlying cause. Accurate measurements of quantities of heat at nanoscale are therefore essential, which fall under the purview of nanocalorimetry. For ease of understanding, let us approach the subject starting from calorimetry and then move toward nanocalorimetry concepts.

What is calorimetry? Derived from the Latin *calor* meaning heat, and the Greek *metry* meaning to measure, calorimetry is the science of measuring the amount of heat of chemical reactions or physical changes. *What is the scope of calorimetry?* Any process that results in heat being generated and exchanged with the environment is a subject for a calorimetric study. *What is a calorimeter?* It is a device used to measure the heat of a reaction. Various types of calorimeters have been designed for operation in different conditions such as adiabatic (thermodynamic process taking place without loss or gain of heat) or isoperibolic (approximately isothermal). Isoperibolic calorimetry is a technique based on calorimetry for measuring thermal effects of chemical reactions or processes, where the temperature is kept constant in the calorimeter jacket.

What is a Calorie? The amount of energy that raises the temperature of 1 mL of water by 1°C.

How does a calorimeter work? A calorimeter is a container with two chambers: (i) one holding the reaction to be measured and (ii) the second chamber containing a measured volume of water. These two chambers are separated

by a metal wall for conducting the heat from the reaction to the water. The temperature of water is measured by a thermometer. The whole assembly is insulated from the environment so that the heat is retained inside the calorimeter. To use the calorimeter, a precisely known amount of pure water is put into the water chamber. The temperature of the water is recorded. Then a precise amount of chemicals is added to the reaction chamber. As the chemical reaction progresses, the temperature rises or falls. The maximum or minimum temperature achieved is noted. The heat of reaction (the amount of heat that must be added or removed during a chemical reaction in order to keep all of the substances present at the same temperature) is calculated in calories by multiplying the difference between final and initial temperatures of water with the volume of water keeping in mind that the density of water is $1\,g\,mL^{-1}$ and its specific heat is $1\,cal\,°C^{-1}$.

What is nanocalorimetry? It is concerned with performing thermal analysis on extremely small amounts of sample (nano- and pico-grams, 10^{-9} and 10^{-12} g). It is a method for measuring heats of reaction of nanoscale samples with nanojoule (10^{-9} J) sensitivity in order to detect reactions and quantify reaction thermodynamics and kinetics (the study of rates of chemical processes) in multilayer structures. Interfaces of interest include those in multilayer advanced gate stacks for microelectronic devices. It is the calorimetry for systems of nanometer dimensions independently formed from deposition, which involves the difficulty in measuring the minuscule amount of heat.

What makes nanocalorimetry interesting? Understanding the thermodynamic properties, e.g., enthalpy (denoted as *H*, thermodynamic quantity equal to the internal energy of a system plus the product of its volume and pressure: $H = E + PV$, where *E* is internal energy, *P* is pressure, and *V* is volume; a measure of the heat content of a chemical or physical system), Gibb's free energy (a thermodynamic property defined to predict whether a process will occur spontaneously at CT and pressure: $G = H - TS$ where *H*, *T*, and *S* are the enthalpy, temperature, and entropy; ΔG is negative for spontaneous processes, positive for non-spontaneous processes, and zero for processes at equilibrium), specific heat, and entropy (*Symbol S*: the quantitative measure of disorder in a system; the amount of thermal energy not available to do work), of nanoscale materials, which are obviously different from their bulk behavior, has become increasingly important as samples with nanometer length scales are routinely produced in the laboratory and semiconductor processing technologies shrink to tens of nanometers in size. Accurate thermodynamic measurements are essential to understand fundamental properties of materials, providing direct insight into the thermodynamics of thin film reactions and phase transitions, the transformation of a thermodynamic system from one phase or state of matter to another. Going forward, new classes of materials may only be synthesized as thin films, a scale at which traditional calorimetric techniques are not useful. The specific heat and other thermodynamic properties of small (nanogram, 10^{-9} g) samples can be quantified, including, ultrathin and multilayered

films, polymer coatings, biomaterials, and nanocrystalline and amorphous materials. Nanocalorimetry can determine the stability of multilayer thin film structures by quantifying the thermodynamics of interfacial reactions. Other applications for nanocalorimeters are the assessment of relaxation in glassy sugars ($C_{12}H_{22}O_{11}$), such as glucose, sucrose, maltose, etc.; evaluation of bulk metallic glass stability; and the quantification of reactions such as nickel-silicide (NiSi) formation in silicon-integrated circuit devices.

One particular phenomenon—particle-size-dependent melting point depression—occurs when the particle size is of the order of nanometers. At these reduced dimensions, the surface-to-volume ratio is high and the surface energy substantially affects the interior bulk properties of the material, e.g., the melting point T_m of nanometer-sized Au particles can be 300 K lower than the bulk value.

Nanocalorimetry research has a great impact on several industries. Performing thermal analysis on smaller and smaller samples can lead to the development of pharmaceutics (the subject that deals with the technology of preparation of medicines) where new material production and purification is a time-limiting step and the amounts of samples are often very small. It can also foster the development of new surface-related technologies that require modification of nanometer-thick surface layers in a controlled manner.

What are the desirable features of a nanocalorimeter? The calorimeter suitable for measuring extremely small amounts of heat should have a very small heat capacity, of comparable magnitude to the nanoscale materials and films in addition to operating over a wide range of temperatures. *How can one get these features?* These features can be obtained by taking advantage of the advanced thin film and membrane fabrication technology because the thermal mass (the ability of a material to absorb and store heat, releasing it slowly) of the calorimeter is dramatically reduced. The low thermal conductance (measure of the amount of heat in calories that will pass through 1 m² of a material of any thickness each second for each degree centigrade difference in temperature on opposite sides of the material) and heat capacity of amorphous Si–N membranes make them ideal for thin film calorimeters both because of their intrinsic properties, such as high Debye temperature (the temperature above which a certain crystal behaves classically, i.e., the temperature above which thermal vibrations are more important than quantum effects) and corresponding low specific heat, and their mechanical strength whereby they can be made extremely thin with a large area. These membranes are made from a low stress, silicon-rich film that is grown by low-pressure chemical vapor deposition (LPCVD). *What is the structure of a nanocalorimeter?* Figure 5.14 shows a schematic and cross section of the nanocalorimeter developed by Queen and Hellman (2009); the sample heaters and thermometers are also shown. *How is the sample heated?* It is heated by the Pt sample heater. The Pt sample heater is used to raise the temperature of the sample area from fractions of a degree to several hundred degrees (300 mK to 800 K) above the frame temperature depending on the measurement technique.
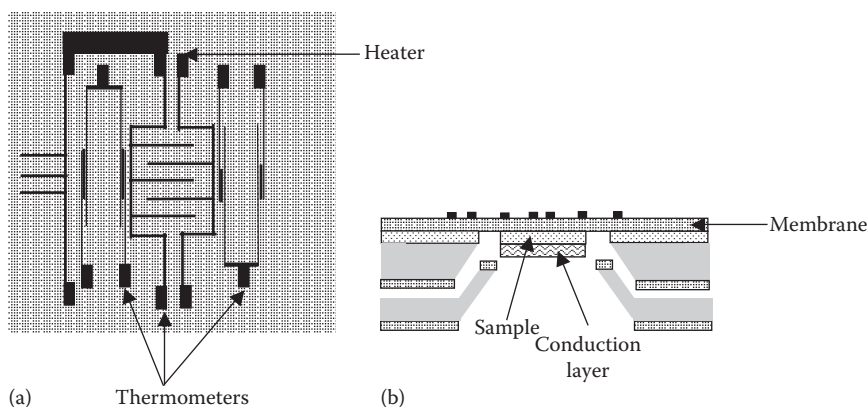
FIGURE 5.14
(a) Geometrical layout and (b) cross-sectional diagram of nanocalorimeter. (After Queen, D. R. and Hellman, F., *Rev. Sci. Instrum.*, 80, 063901-1, 2009.)

How is the temperature of the sample determined? It is measured with the help of three sample thermometers (two a-Nb$_y$Si$_{1-y}$ and one Pt) located on the sample area with impedance-matched thermometers on the frame. The resistance of each thermometer is optimized for a different temperature range. Lead resistances are <2% of the measured resistance for the heater and over the range of each thermometer. The Pt heater and thermometer are 12 μm wide. The a-Nb$_y$Si$_{1-y}$ thermometers have the same doping y and resistances that differ by a factor of 12.5 due to geometry.

How is the heat conducted from the sample to the thermometers? A Cu or Au thermal conduction layer is deposited on the back of the membrane in the same central sample area to ensure that the sample is isothermal (in which the temperature remains constant) with the thermometers.

What are the functions of the different parts of the device? All of the elements on the device are lithographically patterned and their geometries are well known. The 2 mm × 2 mm × 30 nm Si–N membrane acts both as a substrate for the thin film sample and a weak thermal link (a thermal path) between the sample and the Si frame. Samples are deposited onto the backside and in the center of the membrane in the 1 × 1 mm² sample area through a separately micromachined deposition mask.

What are the contributing factors in heat flow? The heat capacities of the membrane, thermal conduction layer, and sample, when present, are the dominant contributions to the total measured heat capacity. The background, or addenda (a list of things to be added), heat capacity of the calorimeter is due to the membrane, conduction layer, and thin film heaters and thermometers. The thermal link that couples the sample area to the environment includes contributions from thermal conduction and radiation. The thermal conduction term depends on the elements that connect the sample area and the frame: Si–N membrane and Pt leads for the heater

and thermometers. Heat loss due to radiation becomes appreciable above 100 K and depends on the membrane area and emissivity (the ratio of the radiant energy emitted by a surface to the radiation emitted by a blackbody at the same temperature).

How is the membrane released? The 30–50 nm thick low stress, amorphous Si–N membrane is released by etching the Si wafer in KOH and then removing the 100 nm oxide layer in buffered HF (prepared by mixing 49% HF and 40% NH_4F in various proportions). The 10–20 nm thick Pt film is lithographically patterned and etched prior to releasing the membrane. a-Nb_ySi_{1-y} is deposited on the wafer with released membranes, patterned, and then etched.

What parameters are decisive for membrane size? The ability to reliably make thin membranes with a large area sets the minimum possible size for the calorimeter. Membranes as thin as 30 nm are sufficiently strong to survive fabrication, processing, and routine handling.

What precautions are taken for stress minimization in deposited films? Residual stress is kept below 200 MPa for all films to reduce strain on the membranes. The stress in the LPCVD Si–N depends on the deposition temperature and ratio of dichlorosilane (SiH_2Cl_2) and ammonia (NH_3) precursor gases. The stress in the sputtered Pt and a-Nb_ySi_{1-y} films is controlled by varying the pressure of the argon (Ar) gas.

How is the device used in heat measurements? The addenda heat capacity of the nanocalorimeter is less than 2×10^{-7} J K^{-1} at room temperature and 2×10^{-10} J K^{-1} at 2.3 K. Heat capacities of several Cu and Au films were measured. The heat capacities of thin Cu and Au films are reported and agree with bulk values. These measurements showed that the nanocalorimeter can be used to measure the heat capacities of films as thin as 30 nm with absolute accuracy <5%, limited by a combination of electrical noise (a random fluctuation in an electrical signal, a characteristic of all electronic circuits), film thickness uncertainty, and ≤2% systematic error (a type of error that deviates by a fixed amount from the true value of measurement; commonly occurs with the measuring instrument having an offset or zero setting error, multiplier, or scale factor error or by the wrong use of instrument or by changes in environment during the experiment) from the measurement technique.

The specific heat and thermal conductivity (the ability of a material to transfer heat, being measured in Watts per square meter of surface area for a temperature gradient of one Kelvin [K] per unit thickness of one meter [W mK^{-1}]) of the thin amorphous silicon Si–N membrane have been extracted and compared with data on thicker membranes. The thermal conductivity of the thin low stress silicon nitride (Si_3N_4) is substantially smaller than thicker membranes, reduced by approximately a factor of 3 over the entire temperature range while the specific heat is enhanced below 20 K and is a factor of 5 at 2 K.

The ongoing development of new nanocalorimeter devices requires advances in thermometry and detailed understanding of the thermal behavior of the materials from which they are made.

5.20 Discussion and Conclusions

Table 5.1 presents the scenario of thermal nanosensors. A look at Table 5.1 reveals the efforts made in different research directions for temperature sensing employing nanotechnology. In one approach, FIB technique was applied to make a small TC W/Pt junction or W resistor. These types of nanosensors are useful for measuring temperatures of nanospots. Utilization of carbon nanowire or CNT for temperature measurements has also been applied. Controllability of carbon nanowire positioning during using FIB writing, and reproducibility of their electrical and structural parameters is an advantage in their favor.

The challenge of the reliable and controllable manipulation with individual CNTs, not met so far, poses a considerable problem for their practical applications. Therefore, in one study, selective growth of CNTs on Ni has been done by CVD. Ag contacts are used. Nickel thin film breaks into small nanoclusters during preheating at 750°C in a CVD furnace due to surface tension as well as the compressive stress due to the mismatch of the thermal expansion coefficients of the float glass substrate and nickel film. This nickel nanocluster acts as a catalyst for growing CNTs in a CVD system. In another study, a thin oxide is deposited on silicon wafer. After this step, a thin film of nickel is deposited and used to make nanoparticles by sputter etch. These nanoparticles are used as catalysts to grow CNT at a later time. Then, a titanium film is deposited and patterned as electrodes and metal leads. CNT is grown in the horizontal direction by MPCVD.

SiNW fabrication process is compatible with top-down complementary metal-oxide semiconductor technology. The wires are patterned through deep UV photolithographic method, etched, and are formed by stress-limited oxidation; deep UV = 224–248 nm. The process uses dry oxidation at 900°C for 2–6 h and subsequent etching away of the silicon dioxide to form the nanowires. The compatibility of this approach with CMOS technology will be an important consideration for mass production of the devices.

The ratiometric NPs are highly temperature dependent in the physiological range (25°C–45°C). Rare earth–doped nanocrystalline fluorescent materials have often been used as temperature sensors because of temperature dependency of their absorption and emission properties. They are particularly useful in many areas where electrical methods are unpractical and are applicable in a wide temperature range (300–900 K). FIR has been determined to measure the temperature. Yb^{3+}–Er^{3+} co-doped nanoparticles have been heated up to 800°C and the temperature has been detected by means of visible upconversion luminescence and visual observation.

CNT accelerometer working on thermal convection principle has also been fabricated. Flow of air adjoining a heated CNT affects its temperature. This principle has been applied to the construction of air-flow sensor using CNT.

TABLE 5.1

Different Types of Thermal Nanosensors

Sl. No.	Name of the Nanosensor	Sensed Quantity/ Applications	Nanomaterial Used	Sensing Principle
1	Nanoscale thermocouple	Thermo EMF	W/Pt	Seebeck effect
2	Tungsten resistive thermal nanosensor	Local thermal measurement	Tungsten	Change of electrical resistivity with temperature
3	Carbon-nanowire-on-diamond resistive temperature nanosensor	Medical and biological applications	Carbon nanowire	The temperature sensitivity of the electrical conductance between the nanowires
4	Carbon nanotube grown on nickel film	Low-temperature sensor (300–10 K)	Carbon nanotubes (CNTs) grown on nickel film deposited on float glass substrate	A very linear metal-like relationship between the electric resistance of CNTs and temperature (TCR ranging from 0.00022 to 0.00064°C^{-1})
5	Laterally grown CNT between two microelectrodes	Good quality temperature sensor	CNT	TCR varies from 0.0008152 to 0.0001759°C^{-1}
6	Silicon nanowire temperature nanosensors: resistance temperature detector (RTD) and Diode temperature detector (DTD)	For micro-reactors	Silicon nanowire	(i) Dependence of electrical conductance of nanowire on temperature. (ii) Variation of reverse-bias diode current with temperature
7	Ratiometric fluorescent nanoparticles	Physiological temperature nanosensors for cellular sensing and imaging	*Temperature indicator*: Eu-DT [(Eu-tris(dinaphthoylmethane)-bis(trioctylphosphineoxide)] and *reference dye*: OASN [N-octyl-4-(3-aminopropyltrimethoxysilane)-1, 8-naphthalimide]	Ratiometric fluorescence between a temperature indicator and reference dye introduced in a nanoparticle

(*continued*)

TABLE 5.1 (Continued)

Different Types of Thermal Nanosensors

Sl. No.	Name of the Nanosensor	Sensed Quantity/ Applications	Nanomaterial Used	Sensing Principle
8	Er^{3+}/Yb^{3+} Co-doped Gd_2O_3 nano-phosphor	Optical thermometry using the green upconversion emission centered at 523 and 548 nm	Combustion synthesized Gd_2O_3: Er_3^+/Yb_3^+ nanocrystalline phosphor	FIR technique. Frequency-upconverted emissions from two thermally coupled excited states $^2H_{11/2}$ and $^4S_{3/2}$ of Er^{3+} centered at 523 and 548 nm in the phosphor pumped by NIR source in temperature range 300–900 K
9	Yb^{3+}–Er^{3+} Co-doped fluoride nanoparticles	In medicine for local hypothermal treatment of cells, for perforation of nanoholes in organics and metals	Rare earth–doped fluoride nanoparticles	Intensity ratio of upconversion luminescence bands
10	Porphyrin-containing copolymer	Full-spectrum colorimeter and ultrasensitive thermometer	ABC triblock copolymer	Color transitions at different thermochromic points
11	Scanning thermal profiler	To provide both temperature and topographic information	Polysilicon–gold thermocouple	A scanning probe microscope with a miniature TC at its tip that provides topographic and thermographic information by sensing heat conducted across a small airgap

12	Superconducting hot-electron nanobolometers	Imaging array radiometers of the terahertz frequency range	Niobium nitride (NbN) thin film	Hot-electron effect
13	Thermal convective accelerometer	Acceleration	CNT bundle	Relation of temperature change of heated CNT convectively affected by acceleration
14	SWCNT sensor for airflow measurement	Airflow	SWCNT	Influence of air flow on heat transference between CNT and its environment
15	Vacuum pressure and flow velocity sensor	Vacuum pressure and air flow	CNT wall	Loss of thermal energy from CNT wall via heat transference through the surrounding gas molecules
16	Nano Pirani gauge	Vacuum pressure	Nano gap between heater and heat sink	Pressure-dependent heat loss from the microbridge to the substrate
17	Carbon nanotube–polymer nanocomposite	Infrared sensor	SWCNT–PC	Enhancement in the IR photoresponse in the electrical conductivity of SWCNTs by embedding the SWCNTs in an insulating polymer
18	Nanocalorimeter	Heat sensor	SIN membrane with sample heaters	Thermal analysis of minute quantities of sample deposited on an ultrathin membrane

The loss of thermal energy from CNT wall has been utilized for vacuum pressure and air flow velocity estimation.

Review Exercises

5.1 Give some examples of carbon-based temperature sensors and point out their relative merits and demerits in terms of flexibility of technology, controllability of positioning of conductors, and reproducibility of electrical and structural parameters.

5.2 How is a pumping at 975 nm of the Yb^{3+}–Er^{3+} co-doped nanoparticles by laser diode resulting in a strong upconversion luminescence signal used as a distant temperature sensing property? Does this cause any heating effect? What are the possible applications?

5.3 In view of penetration capability of terahertz radiation through clothes, dust, smoke, and biological materials being better than infrared or visible light, terahertz imagers are useful for detecting concealed weapons, illicit drugs, and biological materials. Are HEB devices useful for applications requiring low noise temperatures at frequencies from 0.5 to 10 THz? How do they utilize the effect of the electron heating by the incoming terahertz radiation? Discuss the salient features of HEB technology.

5.4 Which sensor provides topographic and thermographic information by sensing heat conducted across a small air gap? Draw and explain the working of a polysilicon–gold TC. Compare its Seebeck coefficient with a tungsten–nickel TC.

5.5 For temperature sensing and imaging, fluorescence-based measurements are useful because of their noninvasive nature, precision, and sturdiness in high-intensity electromagnetic fields. How are fluorescent NPs prepared from a visible-light-sensitized Eu^{3+} chelate used for temperature sensing?

5.6 Are Er^{3+}-doped $BaTiO_3$ nanocrystals suitable for use as FIR-based temperature nanosensor? Is the sensitivity of the nano-thermometer influenced by nonradiative relaxation channels? Do these relaxation mechanisms have any relation to the size of the nanocrystal?

5.7 Describe a temperature sensor fabricated by selectively growing CNTs using the CVD process and discuss its features.

5.8 Discuss the use of Gd_2O_3:Er^{3+}/Yb^{3+} nanocrystalline phosphor as a temperature sensor in the range 300–900 K. Cite some application areas where electrical methods are unpractical.

5.9 MWCNT yarns are available in lengths of 100 m. When embedded in polymers, they would form the basis for a CNT yarn/carbon fiber

composite structure providing local and/or global measurements of strain and temperature. Cite some application examples of such multifunctional strain and temperature sensing.

5.10 Which of the following is a TES: (i) CNT resistor, (ii) superconducting bolometer, (iii) polysilicon–gold TC, and (iv) FIR thermometer.

5.11 From the viewpoint of temperature dependence, which of the following exhibits a metallic as well as a semiconducting nature: (i) carbon nanowire, (ii) CNT, (iii) gold, or (iv) alumina?

5.12 Although the best reported spatial resolution for STP is about an order of magnitude larger than for AFMs and STMs, it permits a wide latitude in the gap size. Explain this with reference to the existence of thermal interaction between the tip and the sample over a much longer range than tunneling current and near-field forces.

5.13 The spatial resolution of infrared thermometry is limited by diffraction to dimensions close to the wavelength of the collected infrared radiation, typically 5 μm at room temperatures. Thermal property variations, temperature gradients, and defects with dimensions smaller than the diffraction limit are inaccessible to far-field infrared thermometry. How does the near-field method help in improving the spatial resolution of infrared thermometry?

5.14 Measurement of phosphor temperature in a cathode ray tube or color TV picture tube in operating condition is of interest for establishing phosphor compositional parameters. Mention one temperature-sensitive mechanism involved in phosphor luminescence that will be useful for temperature measurements.

5.15 What are the thermal and photo effects in the IR photoresponse of the pure $SWNT_{HiPco}$ film? Which effect dominates in the IR photoresponse of $SWNT_{HiPco}$–PC nanocomposites?

References

Agarwal, A., K. Buddharaju, I. K. Lao, N. Singh, N. Balasubramanian, and D. L. Kwong. 2008. Silicon nanowire sensor array using top-down CMOS technology. *Sensors and Actuators A* 145–146: 207–213.

Alencar, M. A. R. C., G. S. Maciel, and C. B. de Araújo. 2004. Er^{3+}-doped $BaTiO_3$ nanocrystals for thermometry: Influence of nanoenvironment on the sensitivity of a fluorescence based temperature sensor. *Applied Physics Letters* 84(23): 4753–4755.

Cherednichenko, S. and V. Drakinskiy. 2008. Low noise hot-electron bolometer mixers for terahertz frequencies. *Journal of Low Temperature and Physics* 151: 575–579, doi: 10.1007/s10909-007-9695-0

Choi, J. and J. Kim. 2010. Batch-processed carbon nanotube wall as pressure and flow sensor. *Nanotechnology* 21: 105502 (8pp.).

Chow, W. W. Y., Y. Qu, W. J. Li, and S. C. H. Tung. 2010. Integrated SWCNT sensors in micro-wind tunnel for air-flow shear-stress measurement. *Microfluid Nanofluid* 8: 631–640.

Chow, L., D. Zhou, and F. Stevie. 2002. Fabrication of nano-scale temperature sensors and heaters. United States Patent 6905736, Filling Date: 02/27/2002, Publication Date: 06/14/2005.

Gaigalas, A. K., L., Li O. Henderson, R. Vogt, J. Barr, G. Marti, J. Weaver, and A. Schwartz. 2001. The development of fluorescence intensity standards. *Journal of Research of the National Institute of Standards and Technology* 106: 381–389.

Gerecht, E. and L. You. 2008. Terahertz imaging and spectroscopy based on hot electron bolometer (HEB) heterodyne detection. *Proceedings of SPIE* 6893, 689308-1-689308-8.

Gianchandani, Y. B. and K. Najafi. 1997. A silicon micromachined scanning thermal profiler with integrated elements for sensing and actuation *IEEE Transactions on Electron Devices* 44(11): 1857–1868.

Jiang, L., S. Shiba, K. Shimbo, N. Sakai, T. Yamakura, M. Sugimura, P. G. Ananthasubramanian, H. Maezawa, Y. Irimajiri, and S. Yamamoto. 2009. Development of THz waveguide NbTiN HEB mixers. *IEEE Transactions on Applied Superconductivity* 19(3): 301–304.

Khosraviani, K. and A. M. Leung. 2009. The nanogap Pirani—A pressure sensor with superior linearity in an atmospheric pressure range. *Journal of Micromechanics and Microengineering* 19: 045007 (8pp.), doi: 10.1088/0960-1317/19/4/045007

Kuo, C. Y., C. L. Chan, C. Gau, C.-W. Liu, S. H. Shiau, and J.-H. Ting. 2007. Nano temperature sensor using selective lateral growth of carbon nanotube between electrodes. *IEEE Transactions on Nanotechnology* 6(1): 63–69.

Ozasa, A., R. Kometani, T. Morita, K. Kondo, K. Kanda, Y. Haruyama, J. Fujita, T. Kaito, and S. Matsui. 2004. Fabrication and evaluation of thermal nano-sensor by focused-ion-beam chemical-vapor-deposition. *Microprocesses and Nanotechnology Conference*, Oct. 27–29, 2004, 28P-7-15: 266–267.

Peng, H.-S., S.-H. Huang, and O. S. Wolfbeis. 2010. Ratiometric fluorescent nanoparticles for sensing temperature. *Journal of Nanoparticle Research*, 12(8): 2729–2733, doi: 10.1007/s11051-010-0046-8

Pradhan, B., K. Setyowati, H. Liu, D. H. Waldeck, and J. Chen. 2008. Carbon nanotube–polymer nanocomposite infrared sensor. *Nano Letters* 8(4): 1142–1146.

Puers, R., S. Reyntjens, and D. De Bruyker. 2002. The NanoPirani: An extremely miniaturized pressure senior fabricated by focused ion beam rapid prototyping. *Sensors and Actuators A: Physical* 97–98: 208–214.

Queen, D. R. and F. Hellman. 2009. Thin film nanocalorimeter for heat capacity measurements of 30 nm films. *Review of Scientific Instruments* 80: 063901-1–063901-7.

Romestain, R., B. Delaet, P. Renaud-Goud, I. Wang, C. Jorel, J.-C. Villegier, and J.-Ph. Poizat. 2004. Fabrication of a superconducting niobium nitride hot electron bolometer for single-photon counting. *New Journal of Physics* 6(129): 1–15.

Saraiya, A., D. Porwal, A. N. Bajpai, N. K. Tripathi, and K. Ram. 2006. Investigation of carbon nanotubes as low temperature sensors. *Synthesis and Reactivity in Inorganic, Metal-Organic and Nano-Metal Chemistry* 36: 163–164.

Singh, S. K., K. Kumar, and S. B. Rai. 2009. Er^{3+}/Yb^{3+} codoped Gd_2O_3 nano-phosphor for optical thermometry. *Sensors and Actuators A: Physical* 149(1): 16–20.

Tikhomirov, V. K., K. Driesen, V. D. Rodriguez, P. Gredin, M. Mortier, and V. V. Moshchalkov. 2009. Optical nanoheater based on the Yb^{3+}–Er^{3+} codoped nanoparticles. *Optics Express* 17(14): 11794–11798.

Vystavkin, A. N., A. G. Kovalenko, S. V. Shitov, O. V. Koryukin, I. A. Kon, A. A. Kuzmin, A. V. Uvarov, and A. S. Il'in. 2010. Hot electron superconducting nanobolometer-sensors for ultrasensitive array radiometers of the terahertz frequency range. *Journal of Communications Technology and Electronics* 55(6): 710–715.

Wei, J., D. Olaya, B. S. Karasik, S. V. Pereverzev, A. V. Sergeev, and M. E. Gershenson. 2008. Ultrasensitive hot-electron nanobolometers for terahertz astrophysics. *Nature Nanotechnology* 3: 496–500.

Yan, Q., J. Yuan, Y. Kang, Z. Cai, L. Zhou, and Y. Yin. 2010. Dual-sensing porphyrin-containing copolymer nanosensor as full-spectrum colorimeter and ultrasensitive thermometer. *Chemical Communications*, 46: 2781–2783, doi: 10.1039/b926882k

Zaitsev, A. M., A. M. Levine, and S. H. Zaidi. 2007. Carbon nanowire-based temperature sensor *Physica Status Solidi* (a) 204(10): 3574–3579.

Zhang, Y., W. J. Li, and O. Tabata. 2009. Extreme-low-power thermal convective accelerometer based on CNT sensing element. In: *The 4th IEEE International Conference on Nano/Micro Engineered and Molecular Systems*, IEEE-NEMS 2009, Shenzhen, China, January 5–8, 2009, pp. 1040–1042.

6

Optical Nanosensors

6.1 Introduction

Optical nanosensors (NSs) are ultra-advanced analytical tools that bring together the advantages of customary sensor technology with flexibility of using dissolved indicators. They are defined as devices with dimensions smaller than 1000 nm that have the capability to incessantly monitor chemical or biological parameters by optically converting the available information into signals useful for analytical purposes. The dimensions of optical NSs vary from a few nanometers, for example, for macromolecules with sensing properties, up to 1 μm. Many micrometer-sized core-shell systems have a sensing layer with a thickness ~ tens to hundreds of nanometers; these are also included in this perspective. These systems often perform analogous to beads having nanometer dimensions and many of them are contrived by identical procedures.

Miniaturization of many fiber-optic sensors to submicron size has been achieved. Moreover, optical fibers impregnated with nanomaterials like carbon nanotubes have been successfully applied. Wide-ranging interest in surface-plasmon resonance (SPR) and localized surface plasmon resonance (LSPR) spectroscopy has been reawakened by the availability of new fabrication methods for plasmonic materials. All these topics will also be included within the scope of this chapter.

Figure 6.1 shows the schematic representation of optical NSs proposed by Borisov and Klimant (2008). They classified NSs into six categories: (i) macromolecular (very large molecules commonly created by some form of polymerization) NSs, (ii) NSs based on polymer materials and sol–gels (gelatinous fluids), (iii) multifunctional core-shell systems, (iv) multifunctional magnetic beads, (v) NSs based on quantum dots (QDs), and (vi) NSs based on metal beads.

Dendrimers are spheroidal (shape of an ellipsoid) or globular (having the shape of a sphere or ball) nanostructures engineered to carry molecules encapsulated in their internal void spaces or attached to their surfaces. The polymeric shell of dendrimers not only imparts water solubility to the indicators, but also protects them against interfering materials and markedly
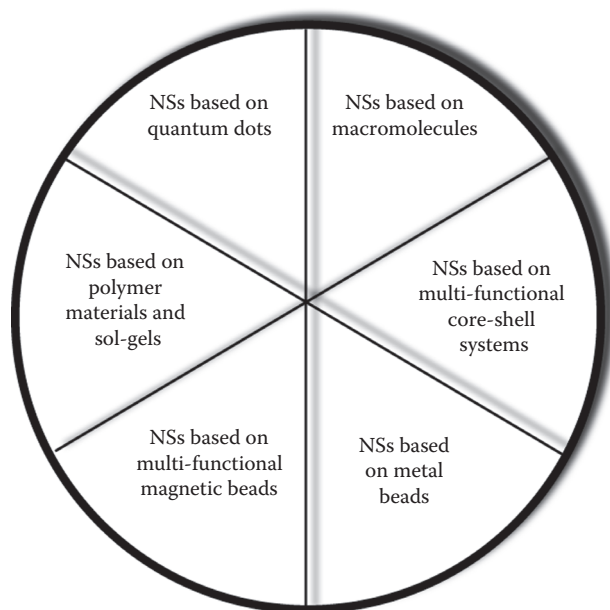
FIGURE 6.1
Different types of optical nanosensors. (After Borisov, S. M. and Klimant, I., *Analyst*, 133, 1302, doi: 10.1039/b805432k, 2008.)

affects the diffusion of analyte molecules, rendering them more similar to other types of NSs than to individual indicator molecules.

Polymer nanoparticles (NPs) generally are formed by microemulsion (an emulsion whose particles are less than 1 μm in size, usually 0.75–2 μm) polymerization, which produces nanosized (~15 nm) particles smaller than those obtainable by emulsion polymerization. Particularly, polar (having a pair of equal and opposite charges) polymers provide water solubility and stability against aggregation. Alternatively, the surface of the hydrophobic (water repellant) beads is modified with polar groups. In contrast to bulk sensor films (typically several microns thick), in NSs many indicator molecules are located near the surface so that leaching (natural process by which water-soluble substances are washed out from a solid) becomes a serious problem, which is avoided with indicators bearing functional groups to be cross-linked with the polymer.

Sol–gels (inorganic silica beads and organically modified silica) are popular materials for designing optical NSs because the beads are easily manufactured by reversed microemulsion polymerization. They are porous, allowing analytes to diffuse freely inside, are robust, and are biocompatible, making them suitable for intracellular use.

NSs based on metal beads are designed on the modification of the surface of a metal nanobead such as Au to render functionality. In contrast to the NSs mentioned earlier, those based on metal beads do not possess an

environment that can significantly tune the sensitivity and selectivity. The main function of the metal core is modulation of fluorescent properties of an indicator located on the surface. Mostly, the metal core acts as a quencher (suppressor of fluorescence), and notably, many such sensors are virtually irreversible.

QDs represent promising alternatives to organic fluorophores (chemical groups responsible for fluorescence). QDs possess good brightness, high photostability (unchanged by the influence of light), broad absorption (radiation enters in a wide range of wavelengths) that allows for their simultaneous excitation, and relatively narrow emission (frequency spectrum of radiation leaving the QD is not wide), which is easily discriminated. Application of QDs in optical sensors can, however, be compromised by a high degree of nonspecific binding and by the fact that many substances act as quenchers (decreasing the fluorescence intensity of a given substance). Susceptibility to quenching often is a serious drawback. However, it can be used to design sensors, providing that other interferences are minimized.

NSs based on other materials include NSs based on single-walled carbon nanotubes (SWCNT), silicon nanowires, and various nanomaterials. The important feature of SWCNTs is their absorption and emission in the near infrared (NIR), the optical window that allows for subcutaneous measurements.

Multifunctional NSs are not only capable of sensing an analyte but also possess additional functionalities that make the material more versatile. Many core-shell materials are multifunctional, for example, the NS for pH in which the core of the bead is stained with an inert fluorescent dye to enable ratiometric referencing. Another type of multifunctional material includes magnetic sensor beads. In contrast to other NSs, magnetic beads can be collected, moved, rotated, or separated from the solution with the help of an external magnetic field. Since they can be collected in the proximity of a detector, the signal-to-noise ratio (a measure of signal strength relative to background noise) is often very high.

As we see, several nano-devices known in the literature as NSs act practically irreversibly and are better classified as "nanoprobes." *What is the difference between NSs and nanoprobes?* Only those systems that respond fully reversibly to the stimulus are designated as "true nanosensors." Nanoprobes do not recover from their altered states. But sometimes the term "nanosensor" is used less exactly and nanoprobes are also considered under NSs.

The optical NSs are extremely diverse in structure, functionality, and materials used in their preparation (Pandana et al. 2008, Zwiller et al. 2008, Couplan et al. 2009). They rely on the same sensing principles as classical optical sensors, namely, dynamic luminescence quenching, Förster resonance energy transfer (FRET), and acid–base equilibria.

What is dynamic luminescence quenching? When an atom or molecule is electronically excited by the absorption of a photon, there are a number of pathways or schemes by which that species can then return to the ground

state. One mode involves the emission of a photon of light, which causes the substance to fluoresce or phosphoresce. Alternatively, species can lose their excitation by nonradiative means: (i) through collisions with other atoms or molecules or (ii) by simply returning to the ground state along a downhill energy path that involves several coupled vibrational and electronic energy states. The first of these nonradiative schemes is called dynamic or collisional luminescence quenching.

What is FRET? It will be elaborately explained in Section 6.11.

What are acid–base equilibria? An acid–base equilibrium system contains an acid and its conjugate base.

The main physical phenomena exploited for optical chemical sensing are absorption, fluorescence, chemical luminescence, Raman scattering, and plasmon resonance. *Optical or light absorption* refers to the process in which energy of light radiation is transferred to a medium. An *absorption* spectrum is the absorption of light as a function of wavelength or frequency. *Fluorescence* is the emission of electromagnetic radiation, especially of visible light, stimulated in a substance by the absorption of incident radiation and persisting only as long as the stimulating radiation is continued. *Chemical luminescence* is the emission of light by a substance caused by *chemical* means. *Raman scattering* is the inelastic scattering of a photon, which creates or annihilates an optical phonon. *Plasmon resonance* is concerned with the excitation of surface plasmons by light.

Optical NSs are favored for nanoscale biological and chemical analysis and environmental applications (Clark et al. 1999, Aylott 2003). Features of optical NSs include the possibility of measuring in very small volumes (including single cells), suitability for three-dimensional imaging, low toxicity, high selectivity, little effect of nonspecific binding and other interferences (such as water) on sensor properties, and versatility in tuning most properties within the polymer matrix. They are divided into two main classes: optical nanobiosensors and optochemical NSs, both of which have become very pervasive and are applied in many fields of science and technology such as biology, biotechnology, environmental analysis, clinical medicine, and marine science, to mention only a few. The instruments, now at advanced stage in labs, demonstrate that optical NSs are some of the best options for scientists and physicians when *in vivo* continuous monitoring is necessary.

As we shall come across the term "biosensor" very frequently in this chapter, let us recapitulate the same. Biosensors are analytical devices for the detection of an analyte that combine a biological component with a physicochemical detector component. They are powerful tools aimed at providing selective identification of chemical compounds at ultra-trace levels in industrial products, chemical substances, environmental samples (e.g., air, soil, and water) or biological systems (e.g., bacteria, virus, or tissue components) for biomedical diagnosis.

6.2 Noble-Metal Nanoparticles with LSPR and UV–Visible Spectroscopy

When the incident photon frequency matches with that of the collective oscillation of conduction electrons, the local electromagnetic fields near the surface of the plasmonic metal film/nanostructure are strengthened (Yonzon et al. 2007). The increase in the electromagnetic fields leads to the intense signals observed in SPR and LSPR. In this section, the NS using LSPR technique is described with reference to its conventional SPR partner bringing out the salient features of LSPR.

For long, the SPR reflectivity measurements have been used to characterize the thickness or refractive index (a constant for a given pair of materials defined as the speed of light in material 1/speed of light in material 2; it measures the speed of light in that substance) of ultrathin organic and biopolymer films at *bulk noble metal* surfaces. Since the introduction of the Biacore (a life science products company, based in Sweden) SPR instrument, uses of SPR spectroscopy have proliferated in the fields of chemistry and biochemistry for characterization of biological surfaces and monitoring binding events. It is a sensor chip–based analytical system where interactions between two or more biomolecules are measured by immobilizing one type of molecule on the surface of a sensor chip and passing a solution containing the other molecule over the surface under controlled conditions. The sensor chip consists of a glass surface coated with a thin layer of gold that provides the physical conditions required for the SPR phenomena.

By monitoring the interaction between the immobilized compound, and one or more compounds in the solution, the technology provides a level of functional data that is difficult or even impossible to obtain by any other single technique. Light does not enter the sample, thereby eliminating the quenching or absorbance (a measure of the light-absorbing ability of an object, expressed as the logarithm to base 10 of the reciprocal of the internal transmittance) problems that beset all spectrophotometric (quantitative measurement of the reflection or transmission properties of a material as a function of wavelength) and fluorescent techniques.

LSPR spectroscopy is a *noble metal NP*–based optical sensing technique, effective for quantitative detection of chemical and biological targets, in which sensing is accomplished by measuring the wavelength shift in the LSPR extinction (state of being extinguished or of ceasing to be) or scattering maximum (λ_{max}) that is induced by the binding of target analytes to the NP surface. The shift in λ_{max} is quantitatively related to the concentration of target analytes. LSPR NSs have been demonstrated as sensitive platforms for the detection of streptavidin, antibiotin, concanavalin (a hemagglutinin, isolated from the meal of the jack bean, which reacts with polyglucosans in the blood of mammals, causing agglutination), Alzheimer's disease

(a progressive, irreversible *disease* characterized by degeneration of the brain cells and commonly leading to severe dementia, a decline in intellectual functioning, including problems with memory, reasoning, and thinking) biomarkers, and many other biorecognition events.

What are the similarities in SPR and LSPR spectroscopy? The NSs based on LSPR spectroscopy operate in a manner totally resembling propagating SPR sensors, which work by transducing small changes in the refractive index near the metallic surface into a measurable wavelength shift response. LSPR NSs induce the small local refractive index changes at the surfaces of metallic NPs. Variations of the reflectivity (the ability of a surface to reflect radiation measured by the fraction of radiant energy that is reflected from it) as a function of the angle of incidence in SPR and extinction peaks (pointed extremities in the plot of *extinction* or transmittance versus wavelength or wavenumber) in LSPR are associated with the same physical phenomenon, that is, collective oscillation of electrons in the metal. The minimum of reflection refers to a maximum of absorption. The behavioral response of SPR and LSPR sensors is described by the following equation (Barbillon et al. 2007):

$$\Delta\lambda_{max} = m\Delta n\left\{1 - \exp\left(-\frac{2d}{l_d}\right)\right\} \tag{6.1}$$

where
  $\Delta\lambda_{max}$ is the wavelength shift
  m is the refractive index sensitivity
  Δn is the change in refractive index induced by an adsorbate (an adsorbed substance)
  d is the effective adsorbate layer thickness
  l_d is the characteristic evanescent (vanishing or likely to disappear) electromagnetic field decay length

This equation is valid for both SPR and LSPR, because they are intrinsically (belonging to the real or essential nature of a thing) related to the same physical phenomenon, albeit applicable to different situations.

What are differences between SPR and LSPR? There are two main differentiating points, namely, the propagating and decay lengths for the two techniques. The propagating length of the surface plasmon is ~10 μm for the SPR and only ~10 nm for LSPR, and the decay length is much smaller for LSPR than for SPR.

This model assumes a single exponential decay (decreasing at a rate proportional to its value, following the exponential function) of the electromagnetic field normal to the planar surface, which is a simplification for the electromagnetic fields associated with metallic NPs. The simplified model enables us to optimize the response of LSPR NSs. The m factor for SPR sensors is about 2×10^6 nm/RIU (refractive index units) and is about 2×10^2 nm/RIU for

LSPR sensors. RIU is the relative change in the index of refraction of the fluid medium. It is often used to describe SPR sensitivity.

The enormous difference in *m* factor values between SPR and LSPR sensors is largely compensated by the very low decay length offered by LSPR gold NP. Indeed this decay length is around 200–300 nm for SPR sensors whereas it is few nanometers for LSPR NSs (for Au NPs l_d = 15 nm). This decay length depends on the size, shape, and composition of the NPs and is responsible for the large sensitivity of the LSPR NS. *What is the importance of this small decay length?* This low decay length allows the detection of a very thin layer of adsorbate molecules.

Even if the sensitivity of the SPR sensors is slightly superior than that of LSPR NSs, a direct comparison is not evident because of the different mechanisms that give rise to their respective sensitivity gains. However, some advantages of LSPR NSs are cited. The LSPR NSs do not require a temperature control compared to SPR since the large refractive index sensitivity of SPR induces a strong dependence on the environmental temperature. No specific angular conditions of excitation are required for LSPR. There is no requirement of prism coupler–based, grating coupler–based, or optical waveguide optical accessories. In practice, SPR sensors require at least $10 \times 10 \, \mu m^2$ area for sensing experiment, whereas for LSPR sensing, the probed zone can be minimized to a large number of individual sensing elements up to a single NP using confocal or near-field measurement techniques. "Confocal" means having the same foci. Confocal microscopy is an optical imaging technique offering several advantages over conventional optical microscopy, including shallow depth of field, and elimination of out-of-focus glare. Resolution and contrast of a micrograph are increased by using point illumination and a spatial pinhole to eliminate out-of-focus light in specimens that are thicker than the focal plane. Near-field measurement techniques include near-field scanning optical microscope in which the intensity of light focused through a pipette with an aperture at its tip is recorded as the tip is moved across the specimen in a raster pattern at a distance of much less than a wavelength; this allows for the surface inspection with high spatial, spectral, and temporal resolving power. Finally, the extinction spectroscopy (highly sensitive optical spectroscopic technique that enables measurement of absolute optical extinction by samples that scatter and absorb light) of LSPR does not need a complex device; an ultraviolet (UV)–visible micro spectrometer is sufficiently efficient to grab the spectra.

Barbillon et al. (2007) used the electron beam lithography (EBL) for the fabrication of these NSs. The EBL system permits to control with high precision the shape, size, and also the distance between the NPs and consequently to tune LSP resonance of metallic NPs arrays in the whole visible range. The obtained metallic nanocylinders have in-plane diameters of 100 and a 200 nm interparticle distance, as viewed by SEM.

They demonstrated the sensitivity of gold NSs by studying the influence of the concentration of 11-mercaptoundecanoic acid (MUA), $HS(CH_2)_{10}CO_2H$,

on the shift of LSPR wavelength. Additionally, to study the selectivity of NSs, the system Biotin/Streptavidin was used to detect very weak concentration of biomolecules. The maximum LSPR wavelength shifts observed for MUA and SA were, respectively, $\Delta\lambda_{max} = 32.8$ and $28\,nm$. They noted that the surface-binding affinity was $(1.1 \pm 0.2) \times 10^3\,M^{-1}$ for the binding of MUA to Au NPs and $(6.1 \pm 0.3) \times 10^9\,M^{-1}$ for the binding between biotin and streptavidin. Furthermore, the limit of detection for the NS was determined to be $1.88 \times 10^{-16}\,mol$ for MUA and $1.25 \times 10^{-18}\,mol$ for SA, corresponding to a density of, respectively, 11300 MUA and 75 SA per NP. The total number of probed molecules was 1.13×10^8 for MUA and 7.5×10^5 for SA, with a probed zone of $30 \times 30\,\mu m^2$.

Most LSPR sensor designs so far only include thin film on a chip or glass slide by conjugating (joined together, especially in a pair; coupled) the metallic NPs with ligands/receptors, that is, containing large ensembles of NP sensors instead of a single NP sensor. The extinction and scattering spectra of plasmonic NPs have spectral shifts that are sensitive to small changes in the local refractive index. Most organic molecules have a higher refractive index than buffer solution. Hence, upon binding to NPs, the local refractive index increases, causing the extinction and scattering spectrum to be red shifted.

Example 6.1

In an LSPR experiment, the observed wavelength shift ($\Delta\lambda_{max}$) was $30\,nm$. The refractive index sensitivity (m) is $2 \times 10^2\,nm/refractive$ index units, the decay length (l_d) is $15\,nm$, and the absorbate thickness (d) is $10\,nm$. Determine the change in refractive index.

Applying Equation 6.1, we have

$$\Delta n = \frac{\Delta\lambda_{max}}{m\left\{1 - \exp\left(-2d/l_d\right)\right\}} = \frac{30}{2 \times 10^2 \left\{1 - \exp\left(-(2 \times 10)/15\right)\right\}}$$

$$= \frac{0.15}{\left\{1 - \exp\left(-1.33\right)\right\}} = 0.204 \text{ refractive index units} \left(\text{RIU}\right) \qquad (6.2)$$

Example 6.2

If one is interested in using LSPR to detect molecular binding taking place on a NS surface, then sensitivity is appropriately expressed in surface coverage unit defined through mass, for example, $pg\,mm^{-2}$. The unit, RU (termed resonance unit or response unit) is defined as 1 $RU = 1\,pg\,mm^{-2}$, and is also often used to determine surface coverage.

If a monolayer of cytochrome c leads to an angular shift of $\sim0.5°$, and the corresponding mass coverage is $\sim3000\,pg\,mm^{-2}$, find the mass sensitivity for an angular sensitivity of 0.1×10^{-3} degree.

0.5 degree angular shift corresponds to mass coverage of $3000\,pg\,mm^{-2}$.

∴1 degree angular shift corresponds to mass coverage of $(3000 \, pg \, mm^{-2})/0.5 = 6000 \, pg \, mm^{-2}$.

∴1×10^{-3} degree angular shift corresponds to mass coverage of $6000 \times 0.1 \times 10^{-3} = 0.6 \, pg \, mm^{-2}$.

Hence, the required mass sensitivity is $0.6 \, pg \, mm^{-2}$ or 0.6 RU.

Example 6.3

Two SPR instruments both use same glass prisms and have similar angular sensitivity, but one uses λ_{low} light and the other uses λ_{high} light where $\lambda_{high} > \lambda_{low}$. Which of these is more sensitive for measuring molecular binding?

The one using λ_{low} light is more sensitive in terms of molecular binding. The penetration length of the evanescence field (typically ~200 nm), created by SPR into the fluid medium increases with the wavelength. Longer wavelengths (e.g., NIR) have the apparent advantage of being able to probe further beyond the sensor surface. However, this results in a significant loss of surface sensitivity. Although longer wavelengths allow for slightly deeper detection into the solution bulk, this results in a significantly lower sensitivity for measuring molecular binding on the sensor surface.

6.3 Nanosensors Based on Surface-Enhanced Raman Scattering

In distinction with the refractive index-based detection schemes, surface-enhanced Raman scattering (SERS) is a vibrational spectroscopic method that yields unique vibrational signatures for small molecule analytes, as well as quantitative information (Jackson and Halas 2004, Kneipp et al. 2010a and b). For applications in biosensors, easily exploitable characteristics of the SERS signal transduction mechanism include their sensitivity, selectivity, low laser power, and no interference from water molecules. Trace analysis of DNA, bacteria, glucose, living cells, posttranslational modification of protein (the chemical modification of a protein after its translation; translation is the third stage of protein biosynthesis in which messenger RNA [mRNA] produced by transcription is decoded by the ribosome to produce a specific amino acid chain, or polypeptide, that will later fold into an active protein) enzymes, chemical warfare agents (asphyxiating or nerve gases, poisons, defoliants, etc.), and carbon nanotubes provide some interesting application examples.

What is the information content of a Raman signal? The Raman spectrum is like a "fingerprint" of a molecule. The Raman scattering signal is composed of many sharp lines. The frequency shift between the excitation light and the Raman lines is determined by the energy of the molecular vibrations $h\nu_M$, which themselves depend on the kinds of atoms and their bond strengths and arrangements in specific molecules.

What is the main advantage of Raman spectroscopy? The main advantage is its capability to provide rich information about the molecular structure of the sample. Sophisticated data analysis techniques based on multivariate (any statistical technique used to analyze data from more than one variable) analysis have enabled exploitation of the full information content of Raman spectra, and inferring about the chemical structure and composition of very complex systems such as biological materials. In this respect, Raman spectroscopy has the potential to become a key tool for health monitoring based on molecular information. The high molecular specificity and large content of chemical information of the technique makes it also a very useful tool for environmental control.

But there is a great disadvantage in any application of Raman spectroscopy. *What is this disadvantage?* It is the extremely small cross section of the effect resulting in very weak spectroscopic signals.

Fortunately, the situation is dramatically improved when a new methodological approach is applied. *What is this new approach?* The new approach combines the interesting optical properties of metal nanostructures with modern laser spectroscopy (in which a laser is used as an intense, monochromatic light source). Due to resonances between the optical fields and the collective motion of the conduction electrons (surface plasmons) in metallic nanostructures, strongly enhanced local optical fields exist in close vicinity of these structures. Exciting opportunities for enhancing spectroscopic signals are available when spectroscopy takes place in these enhanced fields. Raman scattering signals from molecules attached to silver and gold NPs are increased up to 14 orders of magnitude. The effect is called surface-enhanced Raman scattering. It amplifies the Raman scattering cross section (RSCS), an important parameter in the studies and applications of the Raman spectroscopy. It shows the molecular microscopic property as well as the frequency shift and the linewidth. It also indicates the light-scattering capacity of a particular molecule. It is usually obtained by comparing the Raman intensity for an unknown cross section to that for a standard with known cross section.

SERS is an impressive effect for demonstrating the capabilities of this new spectroscopic direction, based on local optical fields. At present, SERS is the only way to detect a single molecule and simultaneously identify its chemical structure.

As a spectroscopic technique, SERS blends the advantages of fluorescence spectroscopy and Raman spectroscopy as illustrated in Figure 6.2. Additionally, since SERS spectroscopy takes place in the local optical fields, the lateral confinement is determined by the confinement of the local fields, allowing the collection of spectroscopic data from volumes below 5 nm dimensions. SERS intensity has been shown to be dependent upon the excitation of the LSPR. To maximize signal strength and ensure reproducibility, it is vital to control all of the factors affecting the LSPR. These factors, including material, size, shape, interparticle spacing, and dielectric environment,
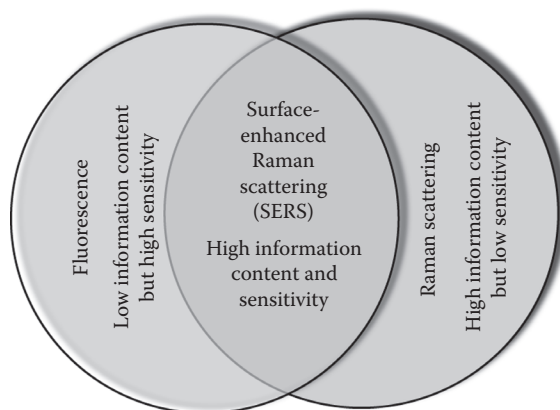
FIGURE 6.2
Coalescence of fluorescence and Raman scattering to form SERS.

must be chosen carefully to ensure that the incident laser light excites the LSPR. The term "enhancement factor" is used to describe the magnitude to which the SERS effect increases the intensity of the Raman scattering for a given experimental system. The enhancement factor is calculated by dividing the SERS spectral intensity by the normal Raman scattering intensity after normalizing both with respect to collection time, laser power, and number of molecules present in the sampling volume.

Figure 6.3 shows the principle of the technology approach for the NS based on SERS. Molecules (unfilled circles) are attached to a gold nanosphere (the big ball). The molecules feel local optical fields, enhanced by factors $A(v)$. In addition to this *electromagnetic field enhancement*, the electronic interaction between the Raman molecule and the metal causes an increase of the Raman cross section itself, known as *chemical or electronic enhancement*. (The filled circles represent atomic scale active sites, where this electronic or chemical interaction occurs.) *What are the relative contributions of chemical and electromagnetically induced enhancements toward the Raman signal?* In many experiments on nanometer-scaled silver or gold structures, the chemical effect provides a contribution of 10–100 to the total SERS enhancement, while the electromagnetic field enhancement gives SERS enhancement factors up to 10–12 orders of magnitude resulting in up to 14 orders of magnitude of net SERS enhancement.

The SERS pH sensor was developed using Ag NPs with Au nanoshell/silica core NPs (Talley et al. 2004). The sensor made of 50–80 nm diameter silver NPs functionalized with para-mercaptobenzoic acid (para-MBA) showed a characteristic SERS spectrum dependent on the pH of the surrounding solution; 4-mercaptobenzoic acid = $(C_7H_6O_2S)$. It was sensitive to pH changes in the range of 6–8. The broad application of SERS technology is greatly hampered by the lack of reliable and reproducible substrates (Netti and Stanford
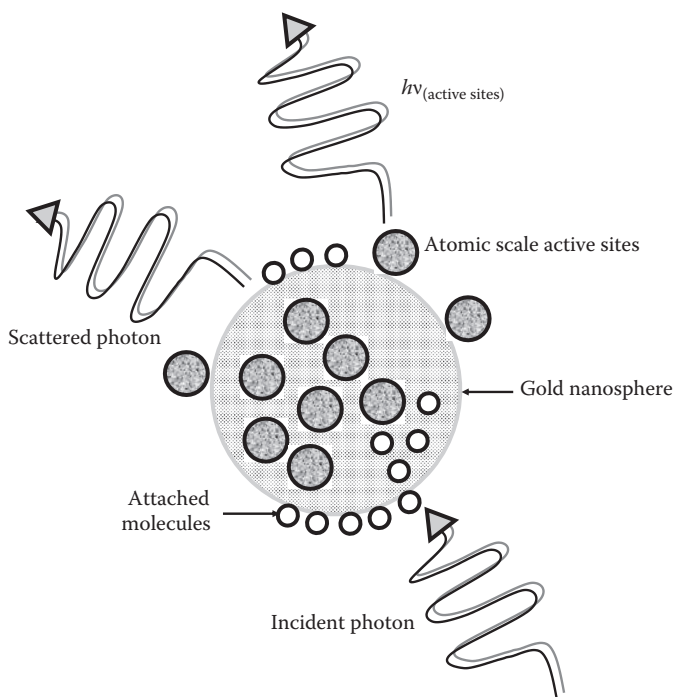
FIGURE 6.3
SERS-based nanosensor.

2006). Cheap and reproducible substrates are required for using SERS as a standard analytical tool. One solution to engineering reproducible SERS substrates is to blend the unique physical properties of photonic crystal devices with the reliable semiconductor manufacturing techniques. Photonic crystal research has established how periodic texturing of surfaces gives rise to new optical properties when the features have dimensions comparable with the wavelength of light ~0.1–1 μm. Metal-coated photonic crystal surfaces result in devices with remarkable dual functionality. They act as *antennae* that concentrate and localize the optical field at the individual features. They also serve as *transducers* that couple the laser beam in and out of the molecule adsorbed on the surface. One advantageous feature of photonic crystal technology is the ability to tune the optical properties of such devices by modifying the dimensions and the geometry of the texture features and the metal. Consequently, the sensitivity of the substrates is tunable to different wavelengths or tailored to the type of experiment.

Additionally, chemical functionality is introduced in the textured gold surface, exploiting well-proven surface chemistry protocols, for example, the coating of SERS substrates with ligand molecules having specific terminal groups is enormously effective in improving the selectivity of the SERS

substrate. This functionalization of SERS substrates is particularly useful for complex molecules such as proteins or for molecules having low affinity to metal.

6.4 Colloidal SPR Colorimetric Gold Nanoparticle Spectrophotometric Sensor

A *colloid* is a substance consisting of particles between 1 and 1000 nm, forming a suspension in a fluid, such as smokes, fogs, foams (a substance that is formed by trapping many gaseous bubbles in a liquid or solid), aerosols (submicron to several micron particles in suspension in the atmosphere), etc., comprising a dispersed phase surrounded by a dispersion medium. A *colorimeter* is a device for measuring color, particularly hue, saturation, and luminous intensity. Hue is one of the main properties of a color, defined technically as the degree to which a stimulus can be described as similar to or different from stimuli that are described as red, green, blue, and yellow (the unique hues). Saturation refers to the dominance of hue in the color.

SPR methods have made far-reaching contribution to the quantification of biomolecular interactions. Disappointingly, conventional SPR reflectometry is difficult to realize in a large-scale array format because of the optics associated with the detection system. This latter limitation is problematical because high-throughput biochemical assays based on protein arrays are necessary to measure the protein–protein and protein–ligand interactions.

Nath and Chilkoti (2002) reported a new label-free optical sensor, which retained most of the desirable features of conventional SPR reflectometry, namely, the ability to monitor the kinetics of biomolecular interactions in real time without a label, but which had several improvements: the sensor was easy to fabricate, and simpler in implementation, requiring only a UV–visible spectrophotometer or flatbed scanner. Importantly, the sensor could be easily multiplexed to enable high-throughput screening of biomolecular interactions in an array-based format.

This optical biosensor exploited colloidal SPR of *immobilized, self-assembled gold NPs on an optically transparent substrate.* The method relies on the change in the absorbance spectrum (the electromagnetic spectrum, broken by a specific pattern of dark lines or bands, presented as a graph of absorbance (*A*) versus frequency or wavelength, observed when radiation traverses a particular absorbing medium; it shows the fraction of incident electromagnetic radiation absorbed by the material over a range of frequencies, a unique property that can be used to identify the material) of a self-assembled monolayer (SAM) of colloidal gold on glass, as a function of biomolecular binding to the surface of the immobilized colloids, Figure 6.4. The implementation of colloidal SPR provides an experimentally simple and convenient biosensor
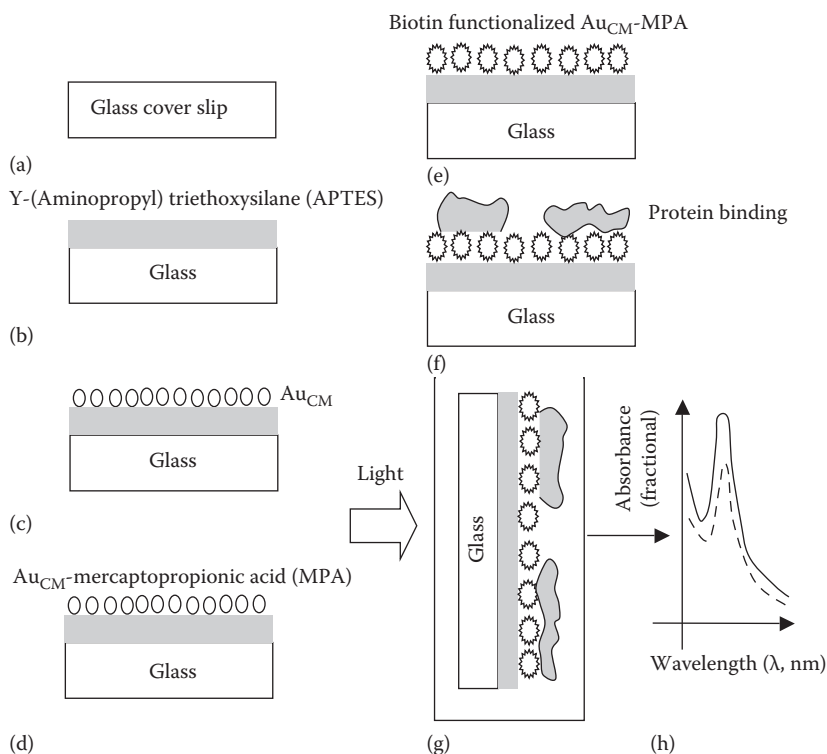
FIGURE 6.4
Steps in the fabrication of colloidal gold sensor ship: (a) Glass substrate. (b) Silanization. (c) Self-assembled monolayer of gold colloids (Au$_{CM}$). (d) Self-assembled monolayer (SAM) of mercaptopropionic acid (MPA). (e) Biotin functionalization. (f) Biomolecular binding. (g) Absorbance study. (h) Shift in absorbance–wavelength curve (dotted curve: before binding and full line: after binding). (After Nath, N. and Chilkoti, A., *Anal. Chem.*, 74, 504, 2002.)

that is easily implemented in most laboratories. The change in the absorbance spectrum associated with biomolecular binding to the immobilized colloids is easily measured in a UV–visible spectrophotometer with an analytical sensitivity and temporal resolution that is adequate to quantify these interactions.

What are the specialties of colloidal SPR? Colloidal SPR is responsible for the intense colors of colloidal solutions of noble metals and is ascribed to the collective oscillations of surface electrons in visible light. It is an interfacial phenomenon and is used in two complementary modes for transduction of binding events at the colloid surface: (i) Changes in the proximity of colloids due to their aggregation in suspension cause a large change in the absorbance spectrum of the colloidal suspension by long-range coupling of surface plasmons. The interparticle distance-dependent color change of colloidal gold due to aggregation of gold colloids has been used in solution-based

immunoassays. (ii) The optical signal arises from the dependence of the peak intensity and position of the surface plasmon absorbance of gold NPs upon the local refractive index of the surrounding medium, which is changed due to binding at the colloid–solution interface. This mode, resembling conventional SPR, has been utilized to determine biomolecular binding on the surface of a colloid in suspension.

Nath and Chilkoti (2002) sought to develop an SPR biosensor in a planar, chip-based format using immobilized gold colloids on an optically transparent substrate because gold colloids permit the transmission of light in the visible region of the electromagnetic spectrum. In principle, a chip-based colloidal SPR sensor enables SPR to be performed in transmission mode in a UV–visible spectrophotometer.

Exploiting the high-affinity of gold colloids for thiol and amine functional groups, a glass surface was transformed into a sensor chip by self-assembly of gold NPs to form a reactive monolayer on an amine-terminated glass substrate. Molecular binding on the sensor surface was transduced to a *colorimetric signal* due to the changes in surface plasmon absorbance of the immobilized gold NPs.

For the experiments shown in Figure 6.4, a glass substrate was functionalized by the formation of an SAM (an organized layer of amphiphilic molecules, molecules having a polar, water-soluble group attached to a nonpolar, water-insoluble hydrocarbon) of aminopropyltriethoxysilane (APTES, $C_9H_{23}NO_3Si$) to present an amine-terminated SAM on the glass surface. Thereafter, the surface was immersed in a solution of colloidal gold, which resulted in the spontaneous self-assembly of a monolayer (CM) of gold colloids on glass (Au CM). The attachment of gold NPs to an amine-terminated surface was strong enough to withstand subsequent chemical modification of the gold NPs without causing their detachment from the surface. The gold colloids were synthesized by trisodium citrate ($Na_3C_6H_5O_7$) reduction of hydrogen tetrachloroaurate ($HAuCl_4$, xH_2O). This is a simple, one-step reduction, and the reaction conditions are controllable to yield monodisperse (monosized) gold NPs of any desired size in the 5–100 nm range.

How is this method advantageous in comparison with conventional SPR? The primary advantage of this sensor is its simplicity and flexibility at several different levels in implementation. First, gold NPs are easily prepared, and can be easily and reproducibly deposited on glass (or other optically transparent substrate) by solution self-assembly. Second, the spontaneous self-assembly of alkanethiols on gold allows convenient fabrication of surfaces with well-defined interfacial properties and reactive groups, which allows the chemistry at the interface to be easily tailored for a specific application of interest, an advantage this sensor shares with conventional SPR on gold or silver films. Third, this sensor enables label-free detection of biomolecular interactions.

The biosensor offers easy multiplexing, enabling high-throughput screening in an array-based format for applications in genomics (a discipline in

genetics concerning the study of the genomes of organisms; an organism's complete set of DNA is called its genome), proteomics (the study of an organism's complete complement of proteins), and drug discovery (the process by which drugs are discovered and/or designed).

Example 6.4

The number of free electrons per unit volume (n) in gold is $5.9 \times 10^{28}\,m^{-3}$. Find the SPR frequency of gold NPs, and the corresponding wavelength in vacuum taking the electron effective mass (m_e) as 10%–15% higher than the free electron mass value. Discuss the effect of the dielectric medium on the obtained values of frequencies and wavelengths.

From Equation 2.26, angular surface plasmon frequency of gold NPs is

$$\omega_M = (q)\sqrt{\frac{n}{3\varepsilon_0 m_e}} \tag{6.3}$$

The concentration of free electrons in gold metal is $5.90 \times 10^{28}\,m^{-3}$. Applying the correction for enhancement of effective mass of electrons in noble metals by 10%–15%, i.e., by 12.5%, the mass becomes $m_e = 9.11 \times 10^{-31}$ $\{1 + (12.5/100)\} = 1.025 \times 10^{-30}\,kg$.

Therefore, angular resonance frequency is

$$\omega_M = (1.6 \times 10^{-19})\sqrt{\frac{5.90 \times 10^{28}}{3 \times 8.854 \times 10^{-12} \times 1.025 \times 10^{-30}}} = 7.45 \times 10^{15}\,rad\,s^{-1} \tag{6.4}$$

and linear resonance frequency is

$$f_M = \frac{\omega_M}{2\pi} = \frac{7.45 \times 10^{15}}{2 \times 3.14} = 1.186 \times 10^{15}\,s^{-1} \tag{6.5}$$

The wavelength is

$$\lambda_M = \frac{c}{f_M} = \left(\frac{3 \times 10^8}{1.186 \times 10^{15}}\right) \times 10^9\,nm = 253\,nm \tag{6.6}$$

where c is the velocity of light in vacuum. The calculated frequency or wavelength in vacuum is in the UV range. Practically, gold NPs display a variety of colors ranging from an intense red color (for particles less than 100 nm) to a dirty yellowish color (for larger particles). The color of a gold NP solution depends on the size of the particles. Small NPs absorb light in the blue–green portion of the spectrum (~400–500 nm) while red light (~700 nm) is reflected, yielding a deep red color. As particle size increases, the wavelength of the SPR-related absorption shifts to longer, redder wavelengths. This means that red light is now adsorbed, and bluer light is reflected, yielding particles with a pale

blue or purple color. As particle size continues to increase toward the bulk limit, SPR wavelengths move into the IR portion of the spectrum and most visible wavelengths are reflected. This gives the NPs clear or translucent color.

Note that the calculation presented earlier is approximate. For gold NPs, the surface plasmon wavelength is 520 nm and the bulk plasmon is at 330 nm.

6.5 Fiber-Optic Nanosensors

Optical fiber–based transduction techniques are very attractive in chemical sensing applications. *What are the reasons that favor optical fibers for sensor applications?* The foremost desirable property is their unique characteristics such as small size, light weight, immunity to electromagnetic interference (the disruption of operation of an electronic device when it is in the vicinity of an electromagnetic field), easy multiplexability, possibility of use in harsh environments, and double function of probe and data communication channel. Another feather in their cap is the versatility of optical sensing because it allows simultaneous collection of intensity and wavelength information and includes a wide range of techniques, for example, absorbance, reflectance, fluorescence, SPR, refractive index, and colorimetry.

Sensor technologies and nanotechnology have joined together offering new opportunities using nanometer-scale structures for chemical and biological applications. Ultrasmall fiber-optic devices show great assurance for sensing applications. Phenomena hitherto below detection levels can now be sensed by nanostructures. Fiber-optic chemical NSs or bio-NSs may provide sensitivity to the single molecule level. Conversely, additional problems may be encountered such as the fundamental challenge of the integration of physical structures and devices on the nanometer scale, as well as their integration into the microscale world. Sensitive coatings at nano- and sub-wavelength-scale based on syndiotactic polystyrene (sPS; thermoplastic semicrystalline material), CNTs, and metal oxides (MOXs) are used in sophisticated fiber-optic chemical NSs.

6.5.1 Fabry–Perot Reflectometric Optochemical Nanosensor Using Optical Fibers and SWCNTs

Diverse optical fiber devices, chiefly standard silica optical fibers (SOFs) and hollow-core optical fibers (HOFs) in the reflectometric configuration and coated long period fiber gratings (LPFGs: the period of the grating considerably exceeds the wavelength of radiation propagating in the fiber), have been deployed for proper integration with chosen sensing overlays. In all cases, the basic procedure adopted derives advantage from the changes in

the optical properties of the sensitive overlays caused by chemical interaction with target analytes to produce modulated light signals. Also, a common feature adopted in all the schemes relies on the use of sensitive overlays at nanoscale. The CNTs possess the outstanding capability for reversible adsorption of molecules of environmental pollutants undergoing a modulation of their electrical, geometrical, and optical properties, such as resistivity, dielectric constant, and thickness.

Penza et al. (2005), Consales et al. (2006, 2007, 2008), and Cusano et al. (2008) worked on the room-temperature sensitivities and response time analysis of SOF sensors, coated with SWCNTs, against aromatic volatile organic compounds (VOCs). SWCNT Langmuir–Blodgett (LB) multilayers were employed as highly sensitive nanomaterials. The LB process is a recognized technique for depositing defect-free, molecularly ordered ultrathin films with prescribed thickness and orientation, allowing fine surface modifications in a multilayered film of carbon nanotubes with a highly controlled manipulation for implementing molecularly self-organizing nanomaterials.

To construct the interferometer, thin films of SWCNTs were transferred on the distal end of a standard silica single-mode optical fiber, Figure 6.5. The
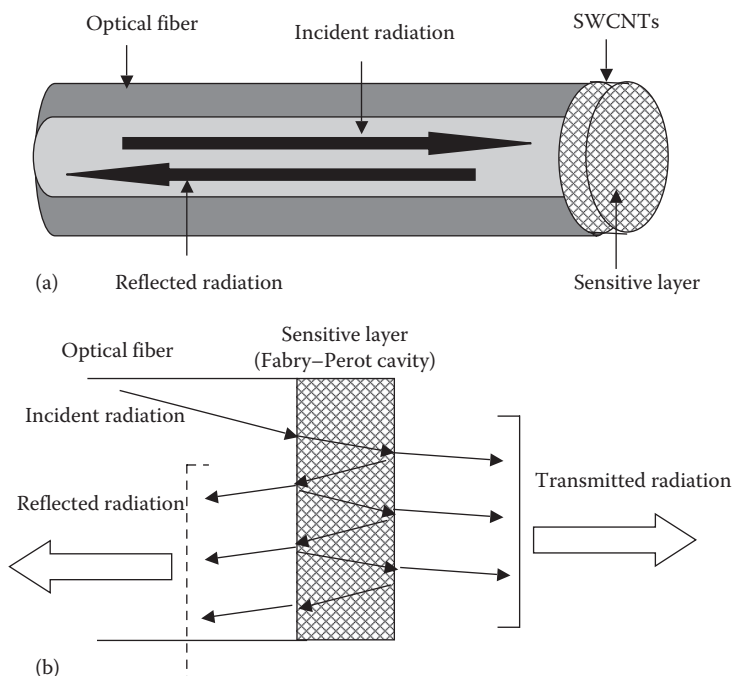
FIGURE 6.5
(a) Optical fiber configuration for vapor testing. (b) Multiple reflected beams in the Fabry–Perot cavity. (After Consales, M. et al., *J. Sens.*, Article ID936074, 29, doi:10.1155/2008/936074, 2008; Cusano, A. et al., *Curr. Anal. Chem.*, 4, 296, 2008.)

term "reflectometric configuration" refers to the standard extrinsic Fabry–Perot (SEFP) configuration working on reflectance measurements from a low finesse (performance) and extrinsic Fabry–Perot interferometer, realized by depositing a flat thin microstructured sensitive film deposited at the distal end of a properly cut and prepared SOF (Cusano et al. 2008). At the fiber–film interface of this arrangement, the light propagating within the optical fiber is partially transmitted and to some extent reflected, depending on the optical and geometrical properties of the SWCNTs overlay. Consequently, any change occurring within the sensitive layer due to chemical adsorption of target analyte molecules, produces a corresponding variation in the film reflectance (R_{SWCNTs}):

$$R_{SWCNTs} = \left| \frac{r_{12} + r_{23} \exp(-i\beta_{SWCNTs})}{1 + r_{12}r_{23} \exp(-i\beta_{SWCNTs})} \right|^2, \quad r_{12} = \frac{\sqrt{\varepsilon_f} - \sqrt{\varepsilon_{SWCNTs}}}{\sqrt{\varepsilon_f} + \sqrt{\varepsilon_{SWCNTs}}},$$

$$r_{23} = \frac{\sqrt{\varepsilon_{SWCNTs}} - \sqrt{\varepsilon_{ext}}}{\sqrt{\varepsilon_{SWCNTs}} + \sqrt{\varepsilon_{ext}}},$$

$$\beta_{SWCNTs} = \frac{2\pi\left(2\sqrt{\varepsilon_{SWCNTs}} \times d_{SWCNTs}\right)}{\lambda} \tag{6.7}$$

where

ε_f and ε_{ext} are the dielectric functions of the optical fiber and external medium

λ is the optical wavelength, and

$$\varepsilon_{SWCNTs} = \varepsilon'_{SWCNTs} - i\varepsilon''_{SWCNTs} \tag{6.8}$$

and d_{SWCNTs} represent the complex dielectric function and the thickness of the SWCNTs layer, respectively.

Upon exposure of the optical probe to external pollutants, the adsorption of pollutant molecules takes place inside the sensitive layer, thereby changing its complex dielectric function and accordingly the optical signal reflected at the fiber–film interface. The sensing property is the dependence of the reflectance at the fiber–sensitive layer interface on the optical properties of the sensitive materials, namely, its complex refractive index ($n^* = n + i \cdot \kappa$; the real part of the refractive index n indicates the phase speed, while the imaginary part κ indicates the amount of absorption loss when the electromagnetic wave propagates through the material. κ is often called the extinction coefficient). The interaction of the field with the chemical molecules present in the environment under investigation occurs not in the volume of the layer but on its surface by means of the evanescent part of

the field, providing a considerable improvement in the sensor performance. The chemo-optic variations induced by the surface–chemicals interaction change the film reflectance and in turn the intensity of the optical signal reflected at the fiber–film interface. The above-mentioned phenomenon has paved the way to a new concept of high-performance chemical sensors based on the manipulation of light through micro- and nanosized structures.

SWCNTs thin films were transferred monolayer by monolayer on to the substrates of the SOF sensors by means of the Langmuir–Blodgett film deposition technique. This deposition technique has been used owing to its capability for depositing ultrathin organic films with precision control over the structure of the film at the molecular level. A solution ($0.2\,mg\,mL^{-1}$) of SWCNTs pristine material in chloroform ($CHCl_3$) was spread onto a subphase constituted by deionized water ($18\,M\Omega$) with $10^{-4}\,M$ of $CdCl_2$. The subphase pH was 6.0 and the temperature 23°C. The monolayer was compressed with a barrier rate of $15\,mm\,min^{-1}$ up to a surface pressure of $45\,mN\,m^{-1}$. The single layer was deposited with a dipping rate of $3\,mm\,min^{-1}$. The transfer ratio of the monolayer from the subphase to the substrate surface was in the range 0.5–0.7. The optical fibers used for the fabrication of the SWCNTs-based optical probes were standard SOFs, having core and cladding diameters of 9 and $125\,\mu m$, respectively. Their effective refractive index was 1.476 at 1310 nm.

In the optical sensor interrogation system (Figure 6.6), reflectance measurements were performed by illuminating the sensing fiber with a pigtailed (with a plait of braided hair) super luminescent light emitting diode (SLED) having a bandwidth of ~40 nm and central wavelength of 1310 nm. For testing and sensitivity comparison of response and recovery times of the optical sensor toward xylene [C_8H_{10}, $C_6H_4(CH_3)_2$ or $C_6H_4C_2H_6$] and toluene (C_7H_8 or $C_6H_5CH_3$) vapors at room temperature, SOF transducers were exposed to the preselected VOC vapors following a time-division multiplexing approach for the parallel interrogation of up to eight optical transducers. Upon target analyte exposures, the SOF sensor showed an increase of the interface reflectance due to the change in the dielectric constant of the sensitive overlay. Optical sensors were more sensitive to xylene than to toluene vapors and provided very high reproducibility and sensitivity with resolution of few hundreds of parts per billion. The effect of the SWCNTs monolayer number on sensor sensitivity and response time was also investigated, showing that the performance of the transducers could be enhanced by tailoring the geometric properties of the sensitive nanomaterial.

6.5.2 In-Fiber Nanocavity Sensor

Elosua et al. (2006) developed a VOCs NS based on Fabry–Perot interferometry using in-fiber nanocavity doped with vapochromic material (showing
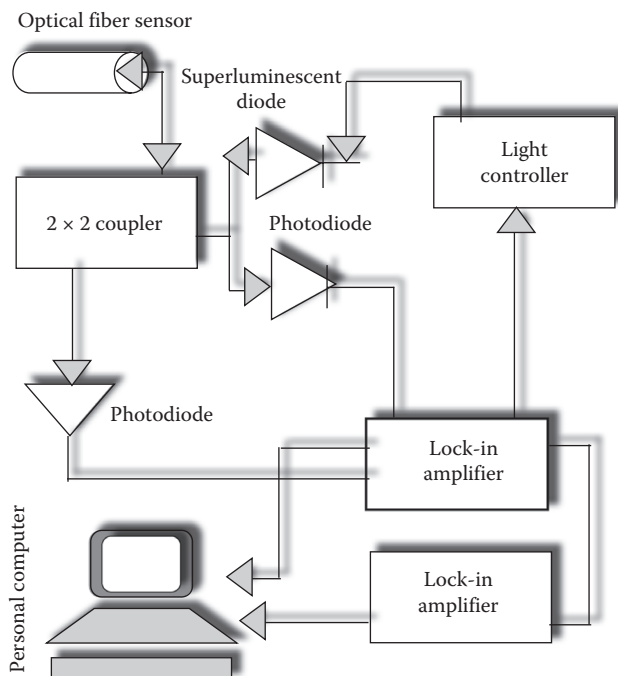
FIGURE 6.6
Optical fiber sensor system. (After Penza, M. et al., *Nanotechnol.*, 16, 2536, 2005.)

pronounced and reversible changes in color and/or emission in the presence of volatile vapors). They prepared a fiber-optic NS (Figure 6.7) by constructing a nanocavity onto a cleaved-ended optical fiber pigtail by electrostatic self-assembly (ESA) method, doping this structure with the vapochromic material belonging to a family of complexes of general formula $[Au_2Ag_2(C_6F_5)_4L_2]_n$, (where L can be pyridine (C_5H_5N), 2,2'bipridine, 1,10 phenantroline ($C_{12}H_8N_2$), ½ diphenylacetylene ($C_{14}H_{10}$), and other ligands) and is able to change its optical properties reversibly in the presence of organic vapors such as ethanol (C_2H_5OH), methanol (CH_3OH), isopropanol (C_3H_8O), acetic acid ($C_2H_4O_2$), and dichloromethane (CH_2Cl_2). The sensor was fabricated using a multimode optical fiber with core and cladding diameters of 62.5 and 125 μm, respectively. Employing a reflection scheme, a change in the intensity modulated reflected signal at 850 nm was found. The interferometric response was produced by the two mirrors of the nanocavity (fiber–film and film–air interfaces). As the bilayers get deposited, the film grows, and, hence, the reflected optical power changes, following an interferometric (two beams) pattern. The nano Fabry–Perot cavity consists of 25 bilayers. The length of the nanocavity was 250 nm (¼th the wavelength used to excite it). The sensor head was able to distinguish among some VOCs and also can determine different concentrations individually.
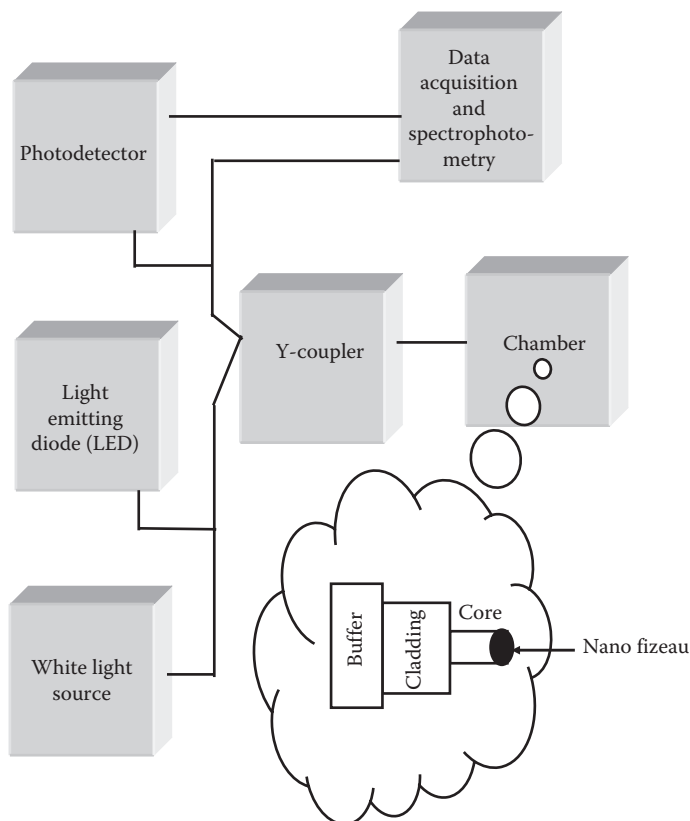
FIGURE 6.7
Block diagram of setup to study variation in optical reflected power on exposure to VOCs and record absorbance spectra. (After Elosua, C. et al., *Sensors*, 6, 578, 2006.)

Changes up to 1.44 dB in the reflected optical power were registered in less than 2 min.

6.5.3 Fiber-Optic Nanosensors for Probing Living Cells

Generally speaking, fiber-optical NSs can be defined as nanometer-scale measurement devices consisting of a layer containing either biological or chemical recognition elements, covalently attached to the transducer. Interaction between the target analyte and the sensing layer produces a physicochemical perturbation that is convertible into a measurable electrical signal. Miniaturization by using optical fibers can make it possible to directly measure biomedical parameters by allowing the probe to be placed inside the body. Fiber-optic nanobiosensors are the tools to investigate biological processes at the cellular level *in vivo*. These NSs have the capability to perform
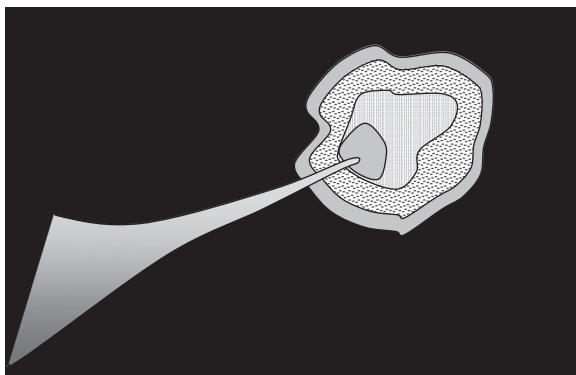
FIGURE 6.8
Fiber optic nanobiosensor inserted into a living cell. Tip diameter of the drawn fiber is 50 nm while that of Ag-coated fiber is 250 nm. (After Vo-Dinh, T. and Kasili, P., *Anal. Bioanal. Chem.*, 382, 918, 2005.)

single-cell analysis (Figure 6.8), which is used for dynamic examination of interactions within individual living cells, which are critical for mapping and deciphering cell-signaling pathways and networks. Not only can antibodies be developed against specific epitopes (sites on the surface of an antigen molecule to which a single antibody molecule binds), but also an array of antibodies can be established to investigate the overall structural architecture of a given protein. Finally, the most noteworthy advantage of NSs for cell monitoring is the *minimal invasiveness* of the technique. The integration of advances in biotechnology and nanotechnology may usher a new generation of nanosystems with unprecedented sensitivity and selectivity for probing sub-compartments of living cells at the molecular level (Vo-Dinh and Kasili 2005). An ideal bioanalytical and biomedical sensor should achieve, in live cells and *in vivo*, real-time tracking of biological/chemical/physical processes as well as detection of disease-related abnormal features, without interferences.

The fabrication of fiber-optic NSs is a crucial precondition for their successful deployment. It involves three steps. First step is the fabrication of nanotip. A fairly well-characterized method for fabricating the nanofiber tips is the so-called "heat and pull technique," involving pulling of nanotips from a larger diameter (600 μm) SOF using a special fiber-pulling device. It is based on local heating of a glass fiber by a CO_2 laser or a heat filament and afterward pulling the fiber apart. The shapes of the resulting tip depend on the temperature and the timing of the procedure. The fiber is then safely put into the fiber-pulling device and the laser-heating source is focused on the median point of the fiber. The optical fiber is pulled to obtain two fibers with submicron tip diameters.

The second step of the nanofabrication process involves coating the tapered side walls of the optical fiber with a thin film of silver, aluminum

or gold (100–200 nm) by thermal vapor deposition, serving to restore the refractive index and enabling propagation of the excitation light down the tapered sides of the nanofiber. Because the fiber tip is pointed away from the metal source, there is no metal coating on the tip. The coating procedure is designed to leave the distal end of the fiber free from silver for subsequent derivatization (a technique used in chemistry that transforms a chemical compound into a product of similar chemical structure, called a derivative), to facilitate covalent immobilization of biorecognition molecules such as antibodies or synthetic peptides (peptides that do not occur in nature, such as those containing unnatural amino acids; peptides are synthesized by coupling the carboxyl group or C-terminus of one amino acid to the amino group or N-terminus of another) coupled to a fluorescent molecular probe, on the exposed silica nanotip. For this purpose, derivatization of the nanotips is done by silanization with glycidoxypropyltrimethoxy silane (GOPS) (3-glycidoxypropyltrimethoxy silane, $C_9H_{20}O_5Si$, is a silane coupling agent (KH-560), a versatile epoxy functional organosilane possessing a reactive organic epoxide and hydrolyzable inorganic methoxysilyl groups) and activation with 1,1'-carbonyldiimidazole (CDI) ((formula $(C_3H_3N_2)_2CO$), an organic compound, white crystalline solid; it is often used for the coupling of amino acids for peptide synthesis). This is the third step. Because the diameter of the optical fiber tip is much less than the wavelength of light used to excite the target analyte, the photons cannot escape from the tip of the fiber by conventional optics. Instead, in a nanobiosensor, after the photons have traversed as far down the fiber as possible, evanescent fields persist traveling through the remainder of the tip, providing excitation for the fluorescent species of interest present in the vicinity of the biosensing layer. Hence, only species that are in extremely close proximity to the fiber tip (i.e., antigens bound to the antibody probes) can be excited, thereby preventing excitation of interfering fluorescent species within other locations of the sample.

Fiber-optic nanobiosensors consisting of antibodies, coupled to an optical transducer element, have been used to detect biochemical targets, benzopyrene tetrol (BPT), and benzo[a]pyrene (BaP), inside single cells; BPT is a metabolite of the carcinogen BaP. *Apoptosis* or programmed cell death is a process by which cells in our tissues and organs degenerate during normal development, aging, or in disease. Fiber-optic nanobiosensors offer a strategy to monitor and measure apoptosis proteins early in the cell death cascade. Furthermore, several different optical fiber–based chemical NSs have been studied for measurement of pH, concentrations of various ions, and other chemical species.

However, as we shall see later, there is some degree of invasiveness still remnant when the fiber pierces the cell wall. Combined with the unavoidable damage associated with readout fiber, there is a little impairment of normal cell functioning and further damage-minimizing techniques are necessary.

6.6 Nanograting-Based Optical Accelerometer

High-precision inertial navigation and seismic sensing for geophysical and oil-field applications require measurement of extremely small acceleration signals (nG/√Hz) at very low frequencies (<100 Hz). Challenges in building such a device were addressed by Krishnamoorthy et al. (2008). These include designing very high sensitivity displacement sensors with low thermomechanical noise and minimizing $1/f$ noise. A practical problem is integrating a large proof mass to the displacement sensor to meet the sensitivity requirements for the device (since larger mass with weak spring constants relates to better sensitivity).

Why is optical method chosen for acceleration measurements? Advantages of optical detection techniques compared to capacitive or piezoresistive methods include high sensitivity and performance close to the Brownian noise (the kind of noise resulting from Brownian motion; hence also called random walk noise) limits of the mechanical structure.

What is a diffraction grating? Diffraction grating is an optical device used to resolve the different wavelengths or colors contained in a beam of light. The device usually consists of thousands of narrow, equidistant, closely spaced parallel slits (or grooves)

An in-plane nanophotonic resonant sensor based on multilayer sub-wavelength optical gratings was used (Carr et al. 2004). Through-wafer holes were incorporated under the nano-grating-based displacement sensors to accommodate transmission detection schemes. This nano-optic sensor comprised two vertically offset layers of sub-wavelength polysilicon nano-gratings separated by an air gap that modulate the near-field intensity and polarization of an incident light source based on the relative lateral motion of the two gratings. The nano-grating-based optical MEMS resonant sensors had extremely high lateral displacement sensitivities (12 fm/√Hz at 1 kHz) and greater than 120 dB of open-loop dynamic range.

Figure 6.9 shows two gratings fabricated to lie on top of each other with a small gap separating the two. The two-layer nano-gratings were fabricated in a surface micromachining process. The wafers underwent a high-aspect-ratio deep reactive ion etch (DRIE) of the substrate silicon to structurally define the proof-mass and the mechanical springs for both the proof-mass and its gimbal (a device consisting of two rings mounted on axes at right angles to each other so that an object, such as a ship's compass, will remain suspended in a horizontal plane between them regardless of any motion of its support). The bottom grating was attached to the compliant mass and moved laterally with it. The upper grating layer was attached to the relatively stiff gimbal.

As the gratings have sub-wavelength dimensions and spacing, they create a local effective index of refraction for that region of space. The index varies as a function of the relative placement of one grating to the other. During an acceleration event, the gratings move relative to each other, modulating the effective index of the structures. This changes the intensity of the reflected
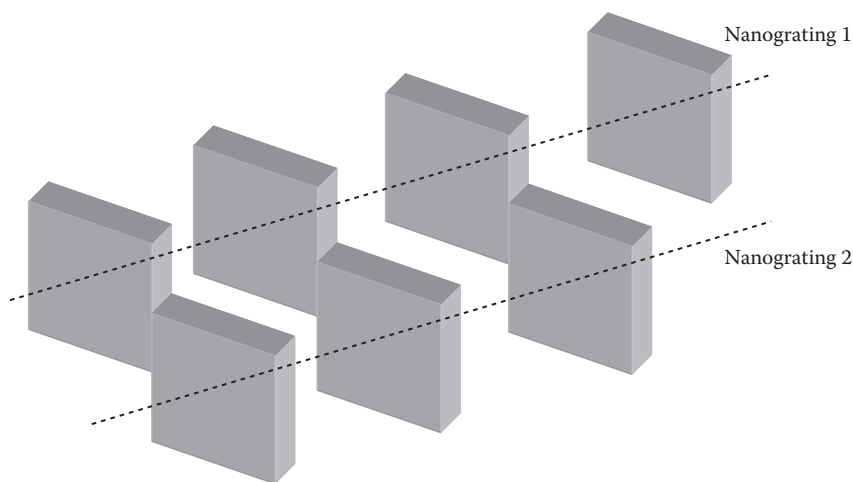
FIGURE 6.9

Nanograting. (After Carr, D. W. et al., Femto-photonics: Optical transducers utilizing novel sub-wavelength dual layer grating structure. In: *Proceedings of the Hilton Head Solid-State Sensors, Actuators, and Microsystems Workshop,* Hilton Head Island, SC, 91, June 2004; Krishnamoorthy, U. et al., *Sens. Actuators, A,* 145–146, 283, 2008.)

and transmitted light. The light typically arrives at normal incident angle on the grating. The reflected intensity is monitored to determine the relative displacement of the gratings, which resulted from accelerations applied to the device. Achieved device sensitivities are $590\,V\,G^{-1}$ and noise floors correspond to $17\,nG/\sqrt{Hz}$ (at 1 Hz); the noise floor is the measure of the signal created from the sum of all the noise sources and undesired signals within a measurement system.

6.7 Fluorescent pH-Sensitive Nanosensors

6.7.1 Renewable Glass Nanopipette with Fluorescent Dye Molecules

Piper et al. (2006) demonstrated the detection of both ratiometric and intensity-based fluorescent changes for physiological levels of pH and sodium by using a simply fabricated fluorescent NS in the form of a 100 nm sized probe. They displayed the concept of the nanopipette as a pH sensor by trapping a negatively charged ratiometric pH-sensitive fluorophore, SNARF-1-dextran, at the pipette tip (Figure 6.10). The sensor was able to measure bath pH from 4.0 to 9.2, responding to the addition of acid in less than 30 ms.

The stability of the fluorescence signal from the NS helped in developing an intensity-based sodium sensor using the negatively charged CoroNa Green dye (Na^+ indicator dye that has its absorbance maximum at 492 nm).
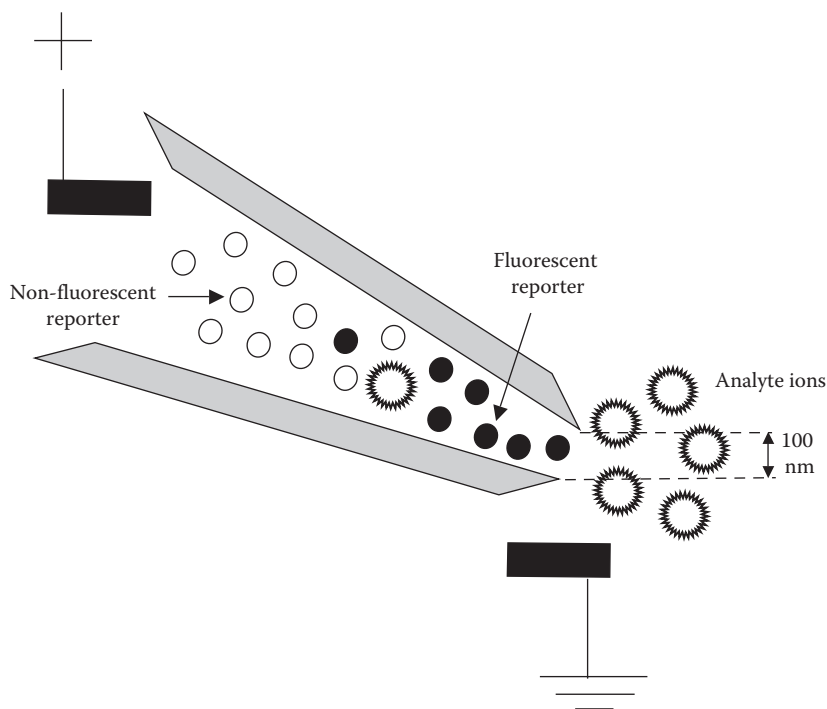
FIGURE 6.10
Glass nanopipette. (After Piper, J. D. et al., *J. Am. Chem. Soc.*, 128, 16462, 2006.)

This dye (50 nM) was trapped at +5 V pipette voltage, and the fluorescence intensity increased as a function of bath NaCl concentration.

The time resolution was measured to be ~2 ms, and the spatial resolution of 600 nm established here can be further improved by bringing the probe and sample nearer to each other. The method is applicable to any fluorescence-based reporter dye and therefore a broad range of analytes and concentrations.

When using free dye molecules, one must concede the problems associated with an uncovered sensing element directly contacting with the intracellular environment and the possible cytotoxic effects this casts upon the biological system. In addition, the effect that the intracellular environment exerts upon the dye is detrimental to acquiring reliable measurements from within the cell. Nonetheless, the small size of the free-sensing dye molecules is a key advantage providing high spatial resolution and allowing information throughout the cell to be collected *en masse*.

6.7.2 Ratiometric pH Nanosensor

Ratiometric measurement uses the ratio of two fluorescent peaks instead of the absolute intensity of one peak. The advantage of the ratiometric sensors

is that factors such as excitation source fluctuations and sensor concentration will not affect the ratio between the fluorescence intensities of the indicator and reference dye. Fluorescent NS particles possess several advantages over direct loading of cells with fluorescent probes, which is a classical method for monitoring metabolic processes of living cells: (i) the biocompatible particle matrix protects the intracellular environment from any toxic dyes, (ii) the cover prevents the dyes from any potential interferences in the cellular environment such as nonspecific binding of proteins, (iii) the nanometer size minimizes the physical perturbation of the cell and the small size provides a fast response time for the sensor, and (iv) several dyes can be included in the same particle.

Sun et al. (2006) reported the synthesis of a prototype ratiometric pH NS. It contains two dyes: a pH-sensitive fluorescent dye and a pH-insensitive reference dye. The dyes are embedded in a polymer matrix, consisting of porous, highly cross-linked polyacrylamide ($-CH_2CHCONH_2-$). Microemulsion polymerization was used for synthesis of particles with diameters ~20–30 nm. To facilitate the insertion of NSs, BY2 tobacco (the most popular and widely used cell line in plant research) protoplasts (plant cells without cell wall) were used as the first model system, and the NS particles were inserted by gene gun bombardment into living protoplasts. Confocal laser scanning microscopy (a tool for obtaining high-resolution images with depth selectivity and 3D reconstructions by acquiring in-focus images from selected depths, a process known as optical sectioning) was used to visualize the fluorescence responses. Responses received from the pH-sensitive dye and the reference dye permitted intracellular pH measurements by fluorescence ratio imaging microscopy. In this method, two different excitation wavelengths are used and the emitted light levels compared.

Leaching of covalently attached dyes from the NPs was found to be minimal when compared to physically embedded dyes. The sensors were stable and robust because (i) the covalent linkage between the dyes and the matrix minimized the leaching of dyes and (ii) the dyes were protected from the environment and vice versa as they are incorporated throughout the matrix. The sensors were capable of reflecting pH in the physiologically relevant range from pH 5.8 to 7.2.

6.7.3 pH-Sensitive Microcapsules with Nanoparticle Incorporation in the Walls

Kreft et al. (2007) introduced an alternative system for pH measurement based on polyelectrolyte microcapsules filled with the pH-sensitive, fluorescent seminaphtho-rhodafluor-1-dye (SNARF1-1) as local pH-probe. Microcapsules are small capsules designed to release their contents when broken by pressure, or when dissolved. They are also useful for drug delivery. *What are polyelectrolytes?* These are polymers of high molecular weight whose repeating units bear an electrolyte group, which dissociate

in aqueous solutions releasing positive and negative ions and making the polymers charged. Naturally, the solutions of such macromolecules in which a substantial portion of the constituent units have ionizable or ionic groups, or both, are electrically conducting like salt solutions and viscous akin to polymers, showing properties similar to electrolytes and polymers.

The fluorophore was retained in picoliter-sized confinement, but, instead of a porous matrix, capsules were used as the carrier system. This capsule-based carrier system allowed for loading fluorophores at two locally distinct sides, the cavity and the walls of the capsules. Furthermore, the capsule wall could be functionalized as demanded by the application, for example, with magnetic or light-sensitive NPs.

SNARF-1 is a fluorescent dye that changes its color of fluorescence upon variations in the hydrogen ion concentration of the surrounding environment. When excited at 488 (or 540) nm, the fluorescence intensities at 580 and 640 nm are related to the pH of the local environment in such a way that the 580 nm intensity decreases with increasing pH, whereas the 640 nm intensity increases with increasing pH values. Thus, the ratio of the intensities of the red and green fluorescence peaks is exploited for pH measurements in the range of pH 6–8. This ratiometric method allows an absolute determination of pH, independent of the amount of sensor molecules, whereas single-wavelength measurements (i.e., by using the pH sensitive dye fluorescein, $C_{20}H_{12}O_5$) would provide only relative determinations of pH. SNARF-loaded capsules changed from red to green fluorescence upon internalization by MDA-MB435S breast cancer cells.

Figure 6.11 shows the steps involved in preparation of microcapsules. One starts with the template (a stencil, pattern or overlay), either positively or negatively charged. The template here is positively charged. The template is made of calcium carbonate $CaCO_3$, polyester, or melamine formaldehyde ($C_4H_6N_6O$) latex particles (*cis*-1,4 polyisoprene or natural rubber that is produced in the milky latex of certain plants; polyisoprene is polymer of isoprene (C_5H_8)). Polyester is a category of polymers that contain the ester functional group in their main chain. They are defined as "long-chain polymers" chemically composed of at least 85% by weight of an ester and a dihydric alcohol and a terephthalic acid [$C_6H_4(COOH)_2$]. Although there are many polyesters, the term "polyester" refers most commonly to a specific material: polyethylene terephthalate (PET), $(C_{10}H_8O_4)_n$. Polyelectrolyte layer of opposite polarity to the positively charged template, the negatively charged layer, is coated on the template by electrostatic attraction. As a result, the particle has an overall negative charge. In sequence (a), succeeding layers of opposite polarities are coated in each step, that is, opposite polarity to preceding step. In the last step of this layer-by-layer (LBL) coating procedure, the template is removed by dissolution. The final charge of the capsule is the charge that was developed in the last step so that the end microcapsule is either positively or negatively charged. Different layers of
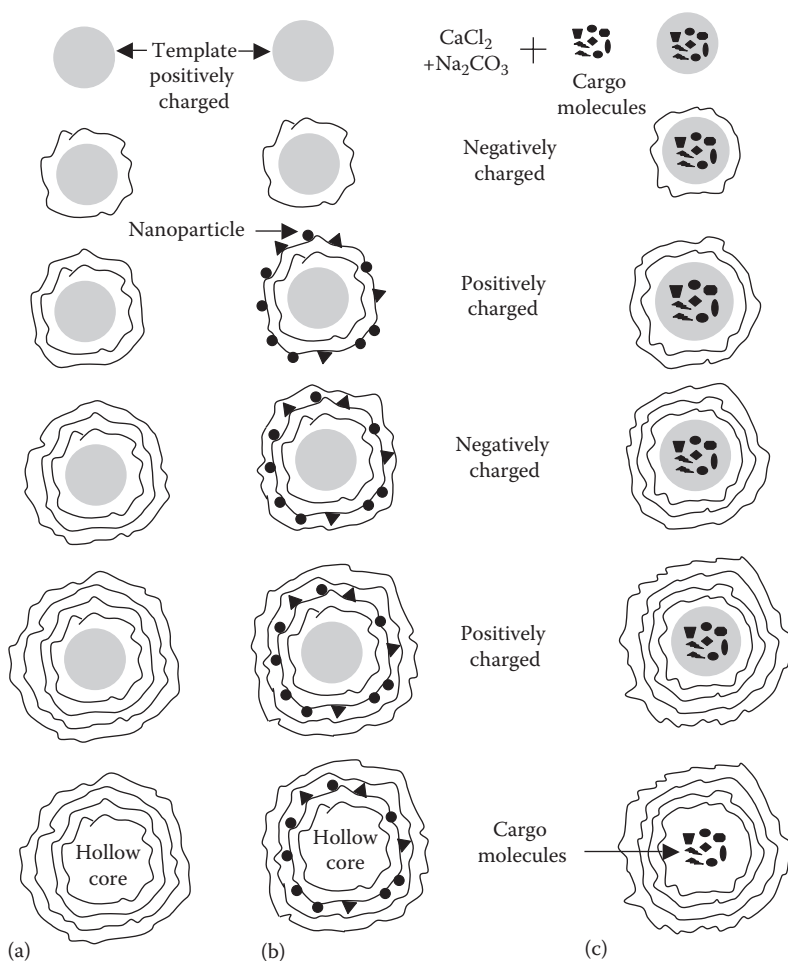
FIGURE 6.11
Steps in the preparation of microcapsules: (a) without nanoparticle incorporation, (b) with nanoparticle incorporation in the side walls, and (c) by embedding cargo molecules inside the core. (After Javier, A. M. et al., In *Biosensing Using Nanomaterials*, Merkoci, A. (Ed.), John Wiley & Sons, New Jersey, 256, 2009.)

the microcapsule are held together by Coulombic attraction. These layers are reminiscent of shells in an onion. The number of layers of polyelectrolyte is decided by the desired final thickness of capsule wall. The capsule wall is a few nanometers thick whereas the overall diameter of the capsule ranges from tens of nanometers to micrometers (the drawing is not to scale). In sequence (b), the intent is to incorporate nanometric particles inside the capsule walls. This is done here after the deposition of first polymer layer. It may be done after any other layer as well, as required. The NPs are covered

by polymer layers and the template is dissolved in the final step. In sequence (c), calcium carbonate is reacted with sodium carbonate to form calcium carbonate molecules, which are mixed with the template particles. The remaining steps are essentially the same but the final microcapsule contains the calcium carbonate cargo molecules that were placed inside the template in the first step. The sequence exemplifies the process for incorporating the required cargo molecules inside the capsule cavity.

6.8 Disadvantages of Optical Fiber and Fluorescent Nanosensors for Living Cell Studies

Optical fiber NSs (optodes) and fluorescent probes suffer from a number of disadvantages for carrying out intracellular measurements. Optical fiber optodes generally consist of a pulled optical fiber with a modified tip, which is inserted directly into the cell. They provide biocompatible platforms that protect both the dye and the cell from each other and avoid adverse interaction between them. The polymer matrix entrapping the chemical-sensing elements also allows for more intricate sensing schemes to be devised due to the close proximity of several elements that can interact complimentarily. All this is achieved at a cost, however, because their large size compared to a single cell causes severe physical perturbations to all but the largest cells. This perturbation arises from the need to collect the signal generated at the tip of the sensor when inserted into the cell, requiring a physical pathway to the detector. Even though optodes have submicron diameters, typical penetration to the point of the nucleus exceeds 1% of the cell volume, inducing severe perturbations and limiting cellular viability. Also, the technique requires highly skilled personnel for properly positioning the optode in the cell and, further, the resultant signals are not spatially resolved. Thus, optodes are bulky and impractical for routine measurements inside living cells due to the connecting fibers, which occupy excessive space inside the cell and cause biological perturbations. Moreover, only low spatial resolution is achievable as very few optodes can be successfully inserted into a cell before it is irreversibly damaged.

Conversely, fluorescent probes, while being easier to use and offering spatially resolved images, still perturb the cell. There are a number of cytotoxic issues concerning fluorescent probes. Variations in signal intensity caused by nonspecific protein binding, and organelle sequestration or efflux of the dye by membrane proteins adversely affect the experiments. They have limited experimental time and once the fluorophore is loaded inside the cell, fluorescent probes often require extensive calibration for individual types of cells.

6.9 PEBBLE Nanosensors to Measure the Intracellular Environment

PEBBLE is the acronym for "photonic explorer or probe for bioanalysis with biologically localized embedding." It is a spherical optical NS, ranging in size between 20 and 200 nm in diameter, transducing chemical or biological events into optical signals. PEBBLE is a generic term covering nano-fabrication techniques that utilize biologically inert polymers to manufacture nanometer-sized spherical optical sensing devices for *in vitro* measurements. PEBBLE NSs derive their origins from the pulled fiber technology, in turn, a development from the optode, an optical sensor device that optically measures a specific substance usually with the aid of a chemical transducer.

PEBBLEs are the tools for biological research realizing the dream of "true silent observers." These NSs have been especially designed to reap the benefits of both the opto-chemical sensors and free dyes while at the same time suppressing their negative characteristics. Functionally, they closely resemble the tip of an optode, as they are essentially a polymer matrix body entrapping chemical sensing elements, such as fluorophores. All the components necessary for carrying the signal out of the cell have been detached, making them more similar to free dyes, but retaining the protective capacity of the biocompatible polymer matrix. Typically only 50 nm in diameter, and with a narrow size distribution, these sensors consume ~1 ppb of a typical mammalian cell causing imperceptibly low physical perturbation. Any chemical perturbation to the cell occurring due to free dyes is annulled, so also the effect of the intracellular environment upon the efficiency of the reporter molecule. In effect, the polymer matrix protects the cell from the dye and the dye from the cell. It has been observed that fluorophores that are adversely affected through random protein binding respond characteristically when entrapped within the polymer matrix of a NS. Succinctly, this NS is a polymeric sphere of nanometer size formed by entrapping chemical sensing elements, typically, fluorophores bound to dextran (any of a group of long-chain polymers of glucose with various molecular weights that are used in confections, in lacquers, as food additives, and as plasma volume expanders).

The polymer matrix is porous, allowing the analytes to diffuse and interact with the entrapped fluorophores, producing a fluorescent response, which is captured by an optical system. Thus, scientists can investigate the intracellular environment in a more natural state by minimizing the chemical and physical effects of free dyes and bulky optodes through the use of polymeric NSs.

PEBBLEs are specifically designed to be minimally invasive, facilitating analyte monitoring in viable single cells or cell cultures (living cells maintained *in vitro* in artificial media of serum and nutrients for the study and growth of certain strains) without perturbing normal cellular functions
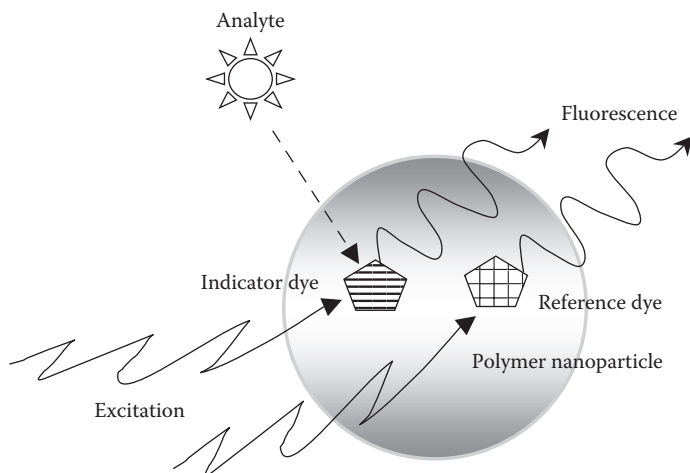
FIGURE 6.12
Polymer nanoparticle sensor using fluorescent indicator and reference dyes. (After Lee, Y.-E. K. et al., *Annu. Rev. Anal. Chem.*, 2, 57, doi: 10.1146/annurev.anchem.1.031207.112823, 2009.)

(Sumner et al. 2006, Lee et al. 2009). PEBBLEs use the same highly selective fluorescent probes normally used in biological monitoring, for example, a probe using Alexa Fluor series of dyes or Oregon Green as reference dyes, is paired with an active fluorescent probe for analyte recognition such as fluo-4 for calcium (fluorescent dye for quantifying cellular Ca^{2+} concentrations in the 100 nM to 1 μM range) or $Ru[dpp(SO_3Na)_2]_3$ [ruthenium(II) tris(4,7-diphenyl-1,10-phenanthroline disulfonic acid) disodium salt] for oxygen, a sulfonated analogue of the well-known oxygen probe (Ru[dpp]3)cl2 (Figure 6.12). The paired probes are encased in a permeable biologically inert matrix, which can be imagined as a sensing platform. The reference fluorophore is chosen specifically to complement the analyte-sensitive probe to ensure that ratiometric analysis is done in either the excitation or emission modes.

The NP sensors based on this design include ions (H^+, Ca^{2+}, Mg^{2+}, Zn^{2+}, Fe^{3+}, and K^+), radicals (OH radical), small molecules (O_2, singlet oxygen, hydrogen peroxide H_2O_2), etc. Singlet oxygen is the lowest excited state of the dioxygen molecule, which is less stable than the normal triplet oxygen. Its lifetime in solution is in the microsecond range (3 μs in water to about 700 μs in C_6D_6). It undergoes several reactions with organic molecules (Ene-Reaction, Diels-Alder Reaction).

The Mg^{2+} PEBBLE sensor is an interesting case exemplifying the design and cellular application. It achieved both high selectivity and sensitivity by using the correct combination of NP matrix and dyes. The traditional measurement of magnesium ion concentrations in biological environments has experienced severe interference from calcium ions. Coumarine 343 is a small hydrophilic dye that is unable to penetrate the cell membrane by itself but is

a very sensitive Mg^{2+} ion probe, having a higher selectivity for magnesium over calcium than any commercially available probe. The Mg^{2+} PEBBLE sensor was constructed by encapsulating this hydrophilic (having an affinity for water; readily absorbing or dissolving in water) dye and a commercial reference dye (Texas Red: $C_{31}H_{29}ClN_2O_7S_2$, sulforhodamine 101 acid chloride, a red fluorescent dye, used in histology for staining cell specimens) in a hydrophilic polyacrylamide (PAM), (formula: $-CH_2CHCONH_2-$, white polyamide, related to acrylic acid) NP. The linear response of these PEBBLEs to Mg^{2+} ion, in the range of 0.1–10 mM, is not affected even in the presence of, simultaneously, 1 mM calcium, 20 mM sodium, and 120 mM potassium, convincingly proving that this Mg^{2+} PEBBLE should serve as a reliable indicator of intracellular Mg^{2+} concentrations. These Mg^{2+} PEBBLEs were utilized to determine the role of Mg^{2+} inside human macrophage cells in the presence of invading salmonella bacteria (a Gram-negative facultative rod-shaped bacteria that are a major cause of food-borne illness throughout the world; it is transmitted to humans through consumption of contaminated food of animal origin, mainly meat, poultry, eggs, and milk) (Martin-Orozco et al. 2006).

Oxygen PEBBLE sensors with enhanced sensitivity and dynamic range were reported using the more sensitive platinum-based oxygen dyes, as well as reference dyes, embedded in a hydrophobic matrix, for example, organically modified silica (ormosil) (Koo et al. 2004) or poly (decyl methacrylate), PDMA (Cao et al. 2004). The hydrophobic matrix is more suitable for oxygen sensing than the hydrophilic one due to the higher oxygen solubility in it.

The embedded platinum(II) octaethylporphine ketone (PtOEPK), another oxygen-sensitive dye, has infrared fluorescence enabling these sensors to work even in human plasma samples, without being influenced by notoriously high light scattering and autofluorescence (background fluorescence) of plasma. The oxygen PEBBLE sensors made of ormosil NPs were applied for real-time imaging of oxygen inside live cells, that is, monitoring metabolic changes inside live C6 glioma (a type of tumor that starts in the brain or spine; it is called a glioma because it arises from glial cells) cells (Koo et al. 2004).

Oxygen sensors have also been prepared by embedding the ruthenium dyes into the polyelectrolyte (an electrolyte of high molecular weight) layers on commercial fluorescent NP surfaces (Guice et al. 2005) or into polymerized phospholipid vesicles (liposomes) (Cheng and Aspinwall 2006). Vesicles are phospholipid-bound globules that usually contain some kind of protein, for example, an enzyme; these store the protein temporarily before they are needed. However, the sensitivities of these sensors were lower being only 60% for polyelectrolyte-coated NP sensors and 76% for polymerized liposome (a synthetic, microscopic spherical sac composed of one or more concentric phospholipid bilayers and used especially to deliver microscopic substances as drugs or DNA to targeted body tissues) sensors.

A singlet oxygen (the diamagnetic form of molecular oxygen O_2) PEBBLE sensor was obtained by incorporating the singlet oxygen–sensitive 9,10-dimethyl anthracene ($C_{16}H_{14}$) or the singlet oxygen–insensitive octaethylporphine ($C_{36}H_{46}N_4$) within ormosil NPs (Cao et al. 2005). The NP matrix enhanced the selectivity of the indicator dye toward singlet oxygen as the matrix blocked the entry of short-lived polar reactive oxidative species (ROS: free radical molecules that contain oxygen in their molecular formula and that are formed during aerobic metabolism), like hydroxyl and superoxide (the superoxide anion with the chemical formula O_2^-) radicals. These nanoprobes were used for monitoring the singlet oxygen produced by "dynamic nanoplatforms (DNPs)" that were developed for photodynamic therapy (a treatment to kill cancer cells involving three key components: photosensitive chemicals, light in the form of controlled and directed laser beams at harmful cells and tissue oxygen) (Reddy et al. 2006). *Nanoplatforms* are nanoscale structures designed as general platforms for multifunctional nanotechnology applications such as nanoporous oxide coatings and superparamagnetic NPs.

A PEBBLE sensor for OH radical was designed by covalently attaching the hydroxyl indicator dye, coumarin-3-carboxylic acid ($C_{10}H_6O_4$), onto the poly(acrylic acid) (PAA), $(C_3H_4O_2)_n$ NP surface, while encapsulating the reference dye deep inside it (King and Kopelman 2003). This design circumvents two potential problems from the hydroxyl radical, the most reactive ROS: (i) inability to penetrate significantly into any matrix without being destroyed and (ii) ability to oxidize (and photobleach: the photochemical destruction of a fluorophore) most potential reference dyes.

How are PEBBLEs fabricated? These sensors are fabricated in a microemulsion (clear, stable, isotropic liquid mixture of oil, water, and surfactant, frequently in combination with a cosurfactant: a thermodynamically or kinetically stable liquid dispersion of an oil phase and a water phase) and consist of fluorescent indicators entrapped in a polyacrylamide ($–CH_2CHCONH_2–$) matrix. A generalized polymerization method is employed permitting production of sensors containing any hydrophilic dye or combination of dyes in the matrix. The PEBBLE matrix protects the fluorescent dye from interference by proteins, allowing reliable *in vivo* calibrations of dyes.

The PEBBLE sensor platform provides a convenient cage that facilitates the entrapment of a number of sensing components, thus offering the ability to fabricate ratiometric sensors from complementary probes or a means to achieve complex sensing schemes with the entrapment of enzymes or ionophores (any of a group of organic compounds that facilitate the transport of ions across the cell, for example, by increasing the permeability of biological membranes). The sensing components are enclosed in a permeable nano-environment, separate from the cellular environment, which offers access only to small analytes while physically preventing interactions of larger proteins with the fluorescent probes. Ratiometric sensors reduce signal variations caused by changes in temperature, focal change shifts, or

photobleaching. Moreover, the fabrication process confers homogeneous distribution of the entrapped dyes within the sensors allowing spatial resolution and continuity to ratiometric measurements in vitro.

How are PEBBLEs introduced into the cells? This is an important issue. A variety of methods to introduce PEBBLEs to a range of cell types have been employed, including functionalization with cell-penetrating peptides (CPPs), endocytosis (cellular process where cells absorb molecules or substances from outside the cell by surrounding and overwhelming them with the cell membrane), commercial lipid transfection (the process of deliberately introducing nucleic acids into cells) agents, cytochalasin D (a mycotoxin: fungal metabolite that blocks cytoplasmic cleavage by blocking formation of contractile microfilament structures resulting in multinucleated cell formation, reversible inhibition of cell movement, and the induction of cellular extrusion; it specifically interferes with actin polymerization), picoinjection (the picoinjector is a robust device to add controlled volumes of reagent using electromicrofluidics at kilohertz rates), and gene gun (a device for injecting cells with genetic information; the payload is an elemental particle of a heavy metal coated with plasmid DNA) bombardment (Webster et al. 2007). The *gene gun* offers high loading of PEBBLE sensors for adherent cells, while *CPP (cell penetrating peptide) functionalization* is effective for cell cultures generally with varying levels of sensor loading, depending on the cell type. *Lipid transfection* gives good results once concentration of transfection reagent and sensor loading are established. *Pico injection* allows location-specific PEBBLE delivery, although the loading is not as high as with other methods and this technique is therefore more suitable for larger cell types where single-cell investigations are of interest. Pinocytosis (a process by which liquid droplets are ingested by living cells by forming narrow channels through its membrane that pinch off into vesicles, and fuse with lysosomes that hydrolyze or break down contents; it is one type of endocytosis) and cytochalasin D cell permeable mycotoxin are unsuitable for preimplantation embryos once the tight junctions have formed, because the space between the cells becomes very small, stopping the entry of nanosensors. Both the method of delivery and cell type of interest are critical to achieve optimal intracellular delivery of NSs.

6.10 Quantum Dots as Fluorescent Labels

Labels or tags constitute the foundations of luminescence and fluorescence imaging and sensing. Labeling or tagging means to incorporate into a compound a substance that is readily detected, such as a radionuclide (an unstable form of a chemical element that radioactively decays, resulting in the emission of nuclear radiation), whereby its metabolism (the chemical processes occurring within a living cell or organism that are necessary for the

maintenance of life) can be followed or its physical distribution detected. The label is the substance thus incorporated.

What types of labels are popular at present? Presently organic dyes (colored substances that have an affinity to the substrate to which they are being applied) are the most versatile molecular labels. But QDs have the potential to substitute organic fluorophores (synthetic molecules such as Oregon Green and Texas Red) in many applications, which have so far been the sole province of these fluorophores. It will be meaningful to study the status of QD applications in this area *vis-à-vis* their organic counterparts. Indeed, one of the most exhilarating, yet also dubious advances in label technology has been the appearance of QDs that have unique optical and chemical properties but complicated surface chemistry, as *in vitro* and *in vivo* fluorophores. The scenario has been comprehensively reviewed by Resch-Genger et al. (2008).

What are the desirable features of a good label? Vital properties of a good label are as follows: (i) A suitable label should be conveniently excitable. Also, the biological matrix should not be concurrently excited and the label should be easily detectable with conventional instrumentation. (ii) It should be bright, that is, having high molar absorption coefficient (an index of how strongly a chemical species absorbs light at a given wavelength) at the excitation wavelength, along with a large fluorescence quantum yield (the ratio of photons absorbed to photons emitted). (iii) It should be soluble in relevant buffer solutions, cell culture media (liquids or gels designed to support the growth of microorganisms or cells), or body fluids. (iv) It must show stability under the conditions of usage. (v) The label should have necessary functional groups for site-specific labeling. (vi) Its photophysics, such as semiconductor band picture of the solid state photophysical properties, should be well known. (vii) Its availability in reproducible quality must be ensured. (viii) Where application demands, its deliverability into cells should be possible without any toxic effects. (ix) Suitability for multiplexing and compatibility for signal amplification are sought-after features.

Typical QDs are core-shell (e.g., CdSe core with a ZnS shell) or core-only (e.g., CdTe) structures functionalized with different coatings. Their properties depend to a considerable extent on particle synthesis and surface modification. Addition of the passivation shell often has the effect of producing a small red shift in absorption and emission relative to the core QD due to tunneling of charge carriers into the shell. In comparison with organic dyes, QDs have the striking property of an absorption that gradually increases toward shorter wavelengths and a narrow emission band of mostly symmetric shape. The spectral position of absorption and emission are tunable by particle size (the so-called *quantum size effect*).

Labeling of biomolecules such as peptides (compounds containing two or more amino acids linked by the carboxyl group of one amino acid to the amino group of another), proteins or oligonucleotides (short nucleic acid polymers) with a fluorophore requires suitable functional groups for covalent

binding or for non-covalent attachment of the fluorophore. The advantage of organic dyes in this aspect is the market availability of a tool-box of functionalized dyes, in conjunction with established labeling protocols. *What is the status about QDs?* There are no consensus methods for labeling biomolecules with QDs. The general principle for QD bio-functionalization is that first the QDs are made water dispersible and then they are bound to biomolecules. Binding is done by choosing one of the several methods like electrostatically, via biotin–avidin interactions or by covalent cross-linking (e.g., carbodiimide-activated coupling between amine and carboxylic groups and maleinimide-catalyzed coupling between amine and sulfhydryl groups and between aldehyde and hydrazide functions) or by binding to polyhistidine tags. *Avidin* is a biotin-binding protein produced in the oviducts of birds, reptiles, and amphibians deposited in the whites of their eggs. *Biotin* is a water-soluble B-complex vitamin (vitamin B_7). A *carbodiimide* or a methane-diimine is a functional group with formula RN=C=NR often used to activate carboxylic acids (oxoacids having the structure RC(=O)OH) toward amide formation; amides are amine (a group of organic compounds of nitrogen, such as ethylamine, $C_2H_5NH_2$, that may be considered ammonia derivatives) derivatives of carboxylic acids. *Maleimide* is an unsaturated imide (a functional group consisting of two carbonyl groups (C=O) bound to nitrogen) with the formula $H_2C_2(CO)_2NH$. A *polyhistidine-tag* is an amino acid motif in proteins that consists of at least five histidine His residues, often at the N- or C-terminus of the protein.

Alternatively, ligands present during synthesis are exchanged for biomolecules containing active groups on the surface. The latter strategy works well for labeling oligonucleotides. A *ligand*, or *complexing agent*, is a polar molecule or ion bonded to a central metal ion.

Currently, only a few standard protocols for labeling biomolecules with QDs are on hand, and the choice of suitable coupling chemistries depends on surface functionalization. It is difficult to prescribe general principles because QD surfaces are unique, depending on their preparation. Accordingly, for users of commercial QDs, knowledge of surface functionalization is essential.

QDs are promising in microarrays (multiplex lab-on-a-chip devices), immunoassays (laboratory techniques that make use of the binding between an antigen and its homologous antibody in order to identify and quantify), and fluorescence *in situ* hybridization (the use of labeled nucleic acid sequence probes for the visualization of specific DNA or RNA sequences). Resolution of single nucleotide polymorphs (any of a group of white blood cells that have lobed nuclei and granular cytoplasm and function as phagocytes by ingesting harmful foreign particles, bacteria, and dead or dying cells) on cDNA arrays has been achieved for mutation in the human oncogene p53. An oncogene is a sequence of DNA that has been altered or mutated from its original form, the proto-oncogene. *Onco-*, from the Greek *onkos*, meaning "bulk," or "mass," refers to the tumor-causing ability of the oncogene. The p53 gene is a tumor-suppressor gene, that is, its activity halts the formation

of tumors. Mutations in p53 are found in most tumor types. A multiplexed analysis for hepatitis B and hepatitis C virus has been done by concurrently using QDs with Plat 566 and 668 nm. "Hepatitis" refers to viral infections that cause inflammation of the liver. Hepatitis A, B, and C are the most common types. Each have different causes and symptoms.

Here, the apparent sensitivity of QD array may be less than observed with organic fluorophores because of their poor excitation efficiency. This shortcoming is overcome by using blue-shifted (a decrease in the wavelength of radiation) source and/or red-shifted (an increase in the wavelength of radiation) LEDs, wherever it is possible.

The ability to track biomolecules within their native environment, that is, on the cell surface or inside the cells, is a key characteristic for any fluorescent label and a prerequisite for assessing molecular function *in vivo. How successful are the QDs in extracelluar (outside the cell) and intracellular (within a cell) labeling?* External labeling of cells with QDs is relatively straightforward, but intracellular delivery is riddled with complexity. Extracellular targeting with QDs is frequently practiced. This is usually accomplished through QD functionalization with specific antibodies to image cell-surface receptors or via biotin ligase–catalyzed biotinylation (the process of covalently attaching biotin to a protein, nucleic acid, or other molecule) in combination with streptavidin functionalized QDs. Whole-cell labeling with QDs has been carried out through microinjection (a simple mechanical process in which a needle roughly 0.5–5 µm in diameter penetrates the cell membrane to insert substances), electroporation (also called electro-permeabilization, is a mechanical method used to introduce polar molecules into a host cell through the cell membrane, using an externally applied electrical field to increase the permeability of the cell plasma membrane; microbiologists have long been applying this method to temporarily punch holes in cell membranes and ferry drugs) or nonspecific or receptor-mediated endocytosis (process by which cells internalize molecules or viruses by engulfing them without passing through the cell membrane; it depends on the interaction of that molecule with a specific binding protein in the cell membrane called a receptor). The labeling specificity and efficiency is improved by using functionalized QDs.

In a milestone study (Dubertret et al. 2002), QDs encapsulated in phospholipid micelles (molecular aggregates of colloidal dimensions) were used to label individual blastomeres (any of the cells produced by cleavage of a fertilized ovum during early development) in xenopus (an African clawed frog) embryos (multicellular diploid eukaryotes in their earliest stage of development). These encapsulated QDs were stable *in vivo*, did not aggregate, and were able to label all cell types in the embryo. At the levels required for fluorescence visualization (2×10^9/cell) the QD-micelles were nontoxic to the cells, but concentrations of 5×10^9/cell produced aberrations (deviations from the normal way). The QDs were restricted to the injected cell and its offspring, though inadvertent translocation to the nucleus was observed at a particular stage in the development of the embryo. Another group labeling

Dictyostelium discoideum (a soil-living amoeba) mentioned that cell labeling was possible for over a week, and that QD labeling had no discernible effects on cell morphology (the form and structure) or physiology (the function). Differently colored QDs could also be used to label various populations in order to scrutinize the effect of starvation on *D. discoideum* (slime mold: a species of soil-living amoeba) development. These cells could be tracked for long duration with no conspicuous fluorescence loss. Zebrafish (a small freshwater tropical fish with barred, zebralike markings) embryo blastomeres (cells resulting from the cleavage of a fertilized ovum during early embryonic development) labeled with QDs and coinjected with CFP, a traditionally used lineage marker (standard cocktail of antibodies designed to remove mature hematopoietic cells from a sample), showed passing of QDs to daughter cells in most cases, although some cells displaying CFP fluorescence displayed no QD fluorescence. This was interpreted to be due to aggregation of QDs, causing unequal inheritance by daughter cells (the cells that result from the reproductive division of one cell during mitosis or meiosis). This is a documented problem, along with fluorescence loss and instability in the QD structure in biological solutions.

What problems arise during intracellular labeling? Generally, owing to their size, the intracellular (existing, occurring, or functioning within a cell) delivery of QDs is not effortless, and compared to organic dyes, the state of the art of delivery of QDs into cells and internal labeling strategies are lagging behind. There is no general procedure to achieve this, and individual solutions need to be empirically established. Several methods have been used to deliver QDs to the cytoplasm (the fluid that fills a cell) for staining of intracellular structures, but are successful to some extent only. Microinjection techniques have been applied to label xenopus and zebrafish embryos, producing pancytoplasmic labeling, but this is a very arduous task, prohibiting high-volume analysis. QD uptake into cells via both endocytic and non-endocytic pathways has also been corroborated. (Endocytosis is the process of cellular ingestion by which the plasma membrane folds inward to bring substances into the cell). But results in only endosomal localization (endosomes are membrane-enclosed vesicles). Two novel approaches have shown pancytoplasmic labeling, by conjugation with Tat protein (transactivator of transcription protein produced by a lentivirus, as HIV, within infected cells that greatly increases the rate of viral transcription and replication and that is also secreted extracellularly where it plays a role in increasing viral replication in newly infected cells and in enhancing the susceptibility of T cells to infection), and by encapsulation in cholesterol (a waxy steroid metabolite)-bearing pullulan (CHP) modified with amine groups; pullulan is a natural water-soluble polysaccharide polymer. Coating with a silica shell may be helpful.

Labeling of specific intracellular structures outside endocytosed vesicles (the vesicle [membrane-bound bubble within the cell] which is formed during the process of endocytosis and which contains substances being imported from outside of the cell. *Endocytosis* is the process by which cells absorb

molecules [such as proteins] by engulfing them) or imaging of cellular reactions in the cytoplasm or the nucleus with QDs necessitates sophisticated tools. Positively charged peptide (a molecule consisting of two or more amino acids)-transduction domains such as Tat (peptide from the cationic domain of the HIV-1 Tat protein), polyarginine (cationic peptide that promotes transfection efficiency under a proper concentration), polylysine (a homopolymer of L-lysine, an essential amino acid), and other specifically designed CPPs can be coated on to QDs for their delivery into cells. The adaptation of other recently developed cell penetrating agents, such as streptaphage, a synthetic ligand based on an *N*-alkyl derivative of 3β-cholesterylamine ($C_{27}H_{47}N$) that efficiently delivers streptavidin to mammalian cells by promoting non-covalent interactions, or polyproline (a sequence of several proline residues in a polypeptide or protein; it tends to generate an anomalous left-handed helix conformation) systems equipped with cationic and hydrophobic (incapable of dissolving in water) moieties (functional groups) for QD delivery is not yet unequivocally confirmed.

How are QDs used for labeling proteins? Labeling of F-actin fibers (filamentous actin; the actin is a protein found in muscle that together with myosin is responsible for muscle contraction) showed that QDs could be used for labeling proteins where preservation of enzyme activity was desirable. Streptavidin-coated QDs were used to label individual isolated biotinylated (being combined with a biotin tag) F-actin fibers. But compared with Alexa 488 (an organic fluorophore), a smaller proportion of labeled filaments were motile (moving or having the power to move spontaneously). Intracellular labeling of these filaments is also practicable. In addition, QDs have been used to label mortalin (a member of the Hsp70 family of proteins), and p-glycoprotein (a protein with covalently attached sugar units), molecules, which are important in tumor cells. Labeling with QDs gave much more photostability than with organic dyes, with a 420-fold increase over Alexa488. This was utilized to image three-dimensionally the localization of p-glycoprotein, with the long fluorescence lifetime allowing successive *z*-sections to be imaged.

What is the application of QDs in multicolor labeling? Owing to the optical properties of organic dyes, their suitability for multicolor signaling at single-wavelength excitation is limited. QDs are the perfect contenders for spectral multiplexing at a single excitation wavelength because of their inimitable flexibility in excitation and their very narrow and symmetric emission bands, which simplify color discrimination. Multiple color labeling of different intracellular structures has been reported (Jamieson et al. 2007). Simultaneous labeling of nuclear structures and actin filaments with QDs of two different colors was done but variable labeling of nuclear structures was observed. Labeling of mitochondria (tubular-shaped organelles acting as cell's power producers) and nuclear structures was done, producing distinct red labeling of the nucleus and green labeling of the mitochondria. Single-color labeling of Her2 (Human epidermal growth factor receptor 2: a gene that sends control signals to cells, directing them to grow, divide, and make

repairs. A healthy breast cell has two copies of the HER2 gene. Some kinds of breast cancer start when a breast cell has more than two copies of that gene, and those copies begin over-producing the HER2 protein so that the affected cells grow and divide much too quickly) has also been shown to be possible and worthy of attention because expression of this is an extrapolative and prognostic (an advance indication or portent of a future event) marker for breast cancer.

What are microenvironmental effects on dyes and QDs? Organic dyes like fluorescein ($C_{20}H_{12}O_5$) and tetramethyl rhodamine isothiocyanate (TRITC), $C_{25}H_{21}N_3O_3S$, and the majority of NIR (NIR) fluorophores have poor photo-stability. For QDs, the microenvironment effect on spectroscopic features is mainly governed by the accessibility of the core surface. This in turn depends on the ligand (and the strength of its binding to the QD surface) and the shell quality. Provided that no ligand desorption occurs, properly shelled QDs are minimally sensitive to microenvironment polarity. Also, QD emission is scarcely responsive to viscosity (a fluid's resistance to flow), contrary to that of many organic dyes, and QDs are less susceptible to aggregation-induced fluorescence quenching (decrease in the fluorescence intensity of a given substance). Limited dye photostability can still hamper microscopic applications requiring high excitation light intensities in the UV–visible light region or requiring long-term imaging. In contrast, adequately surface-passivated QDs display excellent thermal and photochemical (the *effect of light in causing or modifying chemical reactions*) stability. Photo-oxidation (*oxidation* induced by light or some other form of radiant energy) is almost completely suppressed for relevant time intervals as a consequence of their additional inorganic surface layers and shielding of the core material. This is a prominent advantage over organic fluorophores for imaging applications that use intense laser excitation sources or for long-term imaging. Toxicity of organic dyes is not a major problem. In the case of QDs, the cytotoxicity (quality of being poisonous to cells) of elements such as cadmium, which is present in many of these nanocrystals, is well known. Thus, it is critical to recognize whether these cytotoxic substances can leak out of the QD particles over time, upon illumination or oxidation, in addition to whether ligands or coatings are cytotoxic. Cytotoxicity of QDs was observed in some reports in the literature, and in others it was not found. In cases where cytotoxicity was observed, it was usually attributed to leaking of Cd^{2+}, cytotoxic surface ligands, and/or NP aggregation.

Apart from their unique potential for all bioanalytical applications requiring or benefiting from multiplexing, QDs have a bright future in NIR fluorescence *in vivo* imaging, which requires labels that exhibit high fluorescence quantum yields in the 650–900 nm window, have adequate stability, good aqueous solubility, and low cytotoxicity in conjunction with large two-photon action cross sections as desired for deep-tissue imaging. Moreover, QDs are attractive candidates for the development of multifunctional composite materials for the combination of two or more biomedical imaging modalities,

like NIR fluorescence–magnetic resonance imaging. Yet, the routine use of QDs is at present strongly limited by the very small number of commercial systems and the scanty data on their reproducibility and comparability as well as on their potential for quantification.

6.11 Quantum Dot FRET-Based Probes

What is the principle of FRET? FRET involves the nonradiative transfer of fluorescence energy from an excited donor particle to an acceptor particle via dipole–dipole interaction through space. The following conditions must be satisfied by molecular FRET donor–acceptor pairs: (i) overlapping of the fluorescence emission spectrum of the donor and absorption spectrum of the acceptor and (ii) the distance between the donor and the acceptor must be smaller than a critical radius, known as the Förster radius, typically 10–100 Å. This leads to a reduction in the emission and excited state lifetime of the donor, and an increase in the emission intensity of the acceptor.

The rate of transfer of energy, dependent on the spectral overlap between the donor and acceptor, the relative orientation of donor–acceptor transition dipoles and the donor–acceptor distance, is expressed as

$$k_t = \tau_D^{-1} \left(\frac{R_0}{R} \right)^6 \tag{6.9}$$

where

τ_D is the measured lifetime of the donor in the absence of the acceptor
R is the distance between donor and acceptor
R_0 is the Förster distance, the distance at which efficiency of energy transfer is 50%

It is given by the equation

$$R_0 = \left\{ \frac{3000}{4\pi N |A|_{1/2}} \right\}^{1/3} \tag{6.10}$$

where

N is Avogadro's number
$|A|_{1/2}$ is the concentration of acceptor at which the energy transfer efficiency E is 50%

For covalently bound donor–acceptor pair, energy transfer efficiency is written as

$$E = \frac{R_0^6}{R_0^6 + R^6} \qquad (6.11)$$

It is measured experimentally by monitoring the changes in donor and/or acceptor fluorescent intensities or changes in fluorescent lifetimes of the fluorophores; the fluorophore or fluorochrome is a type of fluorescent dye.

FRET is suited to measuring changes in distance, instead of absolute distances, making it appropriate for measuring protein conformational changes (change in shape of a macromolecule in response to its environment or other factors; each possible shape is called a conformation, and a transition between them is called a conformational change), monitoring protein interactions and assaying of enzyme activity (determining enzyme kinetics). Several groups have attempted to use QDs in FRET technologies, particularly when conjugated to biological molecules, including antibodies, for use in immunoassays. These are highly sensitive and specific chemical tests conducted for detecting or quantifying a specific substance, the analyte, in a blood or body fluid sample, using an immunological reaction. Their high specificity arises from the use of antibodies and purified antigens as reagents. An antibody is a protein (immunoglobulin) produced by B-lymphocytes (immune cells) in response to stimulation by an antigen. Immunoassays measure the formation of antibody–antigen complexes and detect them via an indicator reaction.

How are QDs used in FRET? QD-FRET sensors are based on the interactions between QDs serving as donors and molecular fluorophores attached to the QD surface, acting as fluorescent acceptors (Figure 6.13). QDs have been exploited as FRET donors with organic dyes as acceptors, with the QD emission size-tuned to match the absorption band of the acceptor dye. Through FRET mechanism, QDs respond indirectly to environmental changes without any chemical interaction that could affect their photophysical properties and degrade their brightness.

Several QD-based FRET strategies have been developed for detecting nucleic acid, proteins, and other small molecules. Owing to the free choice of the QD excitation wavelength, cross-talk is circumvented in such FRET pairs. However, the distance dependence of FRET means that both the size of the QD itself and that of the surface coating change the FRET efficiency (the percentage of the excitation photons that contribute to FRET); this typically renders QD-FRET less efficient as compared to FRET with organic dyes. Owing to the substantial size of QDs, this limitation can be partly overcome by increasing the number of adjoining small organic acceptor dyes.

Example 6.5

For tryptophan (donor) with dansyl (acceptor), the Förster radius is 2.1 nm. Calculate the FRET efficiency at a distance of (i) 3 nm and (ii) 4 nm.
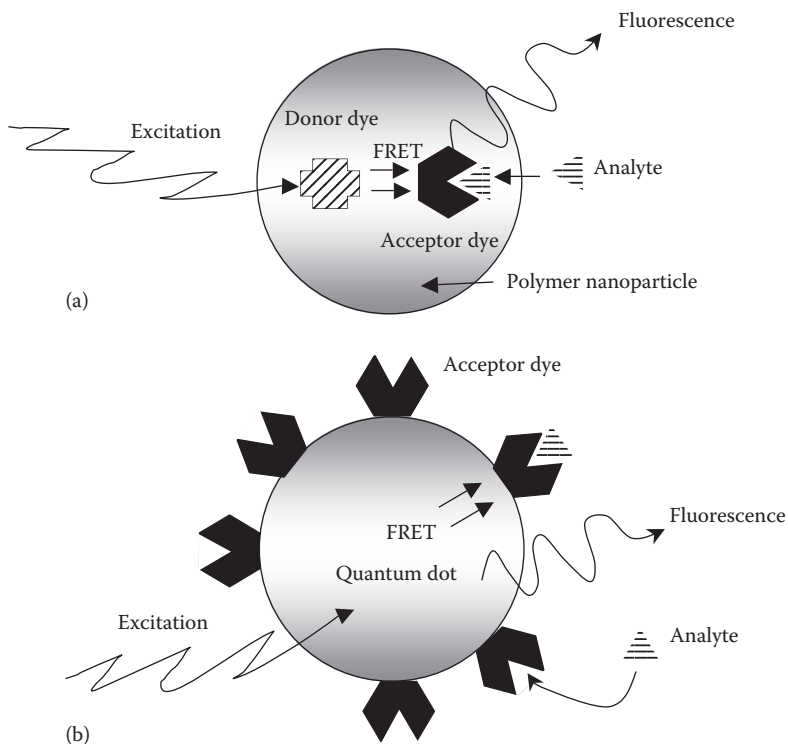
FIGURE 6.13
FRET process using (a) dyes and (b) quantum dot and dye. (After Lee, Y.-E. K. et al., *Annu. Rev. Anal. Chem.*, 2, 57, doi: 10.1146/annurev.anchem.1.031207.112823, 2009.)

Equation 6.11

$$E = \frac{R_0^6}{R_0^6 + R^6} \tag{6.12}$$

gives

(i) $\quad E = \dfrac{(2.1)^6}{2.1^6 + 3^6} \times 100\% = \dfrac{85.766}{85.766 + 729} \times 100\% = 10.53\% \quad$ at $R = 3\,\text{nm}$

$$\tag{6.13}$$

and

(ii) $E = \dfrac{(2.1)^6}{2.1^6 + 4^6} \times 100\% = \dfrac{85.766}{85.766 + 4096} \times 100\% = 2.051\% \quad$ at $R = 4\,\text{nm}$

$$\tag{6.14}$$

Example 6.6

In a FRET experiment, the Förster radius R_0 was 2 nm and the fluorescence lifetime was 5×10^{-9} s. What is the rate of energy transfer at (a) $R = 10$ nm and (b) 5 nm? Comment on the variation.

(a) From Equation 6.9 at $R = 10$ nm,

$$k_t = \tau_D^{-1} \left(\frac{R_0}{R} \right)^6 = \frac{1}{5 \times 10^{-9}} \left(\frac{2}{10} \right)^6 = \frac{6.4 \times 10^{-5}}{5 \times 10^{-9}} = 1.28 \times 10^4 \text{s}^{-1} \quad (6.15)$$

(b) At $R = 5$ nm,

$$k_t = \tau_D^{-1} \left(\frac{R_0}{R} \right)^6 = \frac{1}{5 \times 10^{-9}} \left(\frac{2}{5} \right)^6 = \frac{4.096 \times 10^{-3}}{5 \times 10^{-9}} = 8.192 \times 10^5 \quad (6.16)$$

As R decreases from 10 to 5 nm by a factor of 2, the rate of energy transfer increases by a factor of $8.192 \times 10^5 / 1.28 \times 10^4 = 64$.

6.11.1 QD-FRET Protein Sensor

QD-FRET has been used for monitoring protein interactions in the Holliday Junction (McKinney et al. 2003, Hohng and Ha 2005, Jamieson et al. 2007). The Holliday Junction is an intermediate in the recombination of DNA that undergoes conformational change on addition of Mg^{2+} ions. QD585 was used as a donor on one arm of the DNA, and Cy5 as an acceptor on a perpendicular arm. Movement of the arms on addition of Mg^{2+} was detected as a change in the emission of both donor and acceptor. However, the changes were detected with noticeably less efficiency than with the equivalent Cy3/Cy5 FRET. Cy3 and Cy5 are reactive, water-soluble fluorescent dyes of the cyanine dye family. Cy3 dyes are red (~550 nm excitation, ~570 nm emission, and appear red), while Cy5 is fluorescent in the far red region (~650/670 nm) but absorbs in the orange region (~649 nm). These dyes are small organic molecules, and are typically conjugated to proteins via primary amines, that is, lysines.

Generally, the disadvantages of organic dyes and fluorescent proteins for FRET applications have their foundations in cross talk. This cross talk results from direct acceptor excitation due to the relatively broad absorption bands of these fluorophores. Additional difficulties are met in spectral discrimination of the fluorescence emission, owing to their relatively broad emission bands, their small Stokes shifts (the difference (in wavelength or frequency units) between positions of the band maxima of the absorption and emission spectra, for example, fluorescence and Raman of the same electronic transition), and the "red tails" of the emission spectra in the case of dyes like fluoresceins ($C_{20}H_{12}O_5$, yellow crystals), rhodamines (red fluorescent dyes: supplements to fluoresceins, as they offer longer wavelength emission

maxima and provide opportunities for multicolor labeling or staining), *boron-dipyrromethene* (BODIPY, a class of fluorescent dyes) and cyanines. This renders tedious correction of measured signals mandatory.

6.11.2 QD-FRET Protease Sensor

Shi et al. (2006) synthesized a QD-FRET-based protease (enzyme that catalyzes the hydrolytic breakdown of proteins) sensor to measure the protease activities *in vivo*. QDs coated with unlabeled RGDC (R: Purine, G: Guanine, D: not C, and C: Cytosine) peptide (Organic substance in which the molecules composed of two or more amino acids joined through amide formation, are structurally like those of proteins, but smaller; includes many hormones, antibiotics, and other compounds that participate in the metabolic functions of living organisms) molecules emit green light. When capped with rhodamine-labeled RGDC molecules, the emission color changes to orange through FRET interactions between the QDs and rhodamine molecules (serving as acceptors). Upon enzymatic cleavage of the RGDC peptide molecules, the rhodamine (acceptor) molecules no longer provided an efficient energy transfer channel to the QDs. Consequently, the emission color of the QDs changed back to original green (Figure 6.14).

Through a unique cysteine ($C_3H_7N_1O_2S_1$), approximately 48 copies of a rhodamine labeled collagenase (enzyme that breaks the peptide bonds in collagen, a fibrous protein used to connect and support other body tissues) substrate RGDC were coated on each QD (Zhou and Ghosh 2006). The emission wavelength of QD at 545 nm matched the absorption wavelength of rhodamine. Upon peptide cleavage by collagenase, QD fluorescence at 545 nm increased and rhodamine fluorescence at 590 nm decreased. Ratiometric changes in fluorescence were employed to quantitative study of protease activity in HTB126 cancer cell lines (human mammary gland; disease: carcinoma) in which collagenase is aberrantly expressed. These experiments showed that real-time measurements of protease activity were possible by using QD-FRET-based ratiometric sensors. Since activities of different proteases are measurable by changing the peptide sequences, more proteolytic enzyme (enzyme that breaks the long chainlike molecules of proteins into shorter fragments, peptides, and eventually into their components, amino acids) assays are likely to be performed utilizing QD-FRET-based sensors Further, there are opportunities for the introduction of multiplexing (sending multiple signals or streams of information on a carrier at the same time in the form of a single, complex signal and then recovering them) and high-throughput formats.

6.11.3 QD-FRET Maltose Sensor

As QDs have broad absorption spectra and narrow emission spectra, they are usually used as fluorescence donors. The fluorescence acceptor is often an appropriately labeled peptide (any of various natural or synthetic
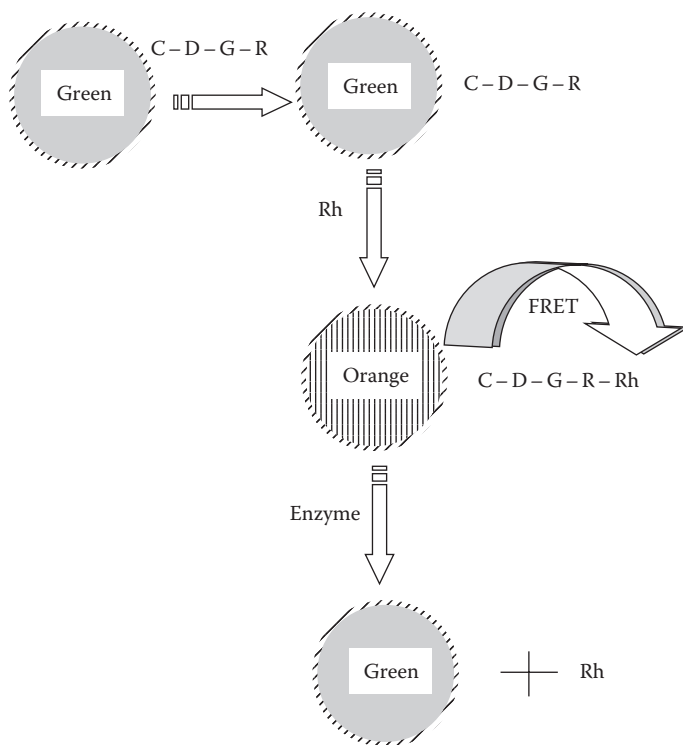
FIGURE 6.14
Scheme of QD protease biosensor. (After Shi, L., et al., *J. Am. Chem. Soc.*, 128, 10378, 2006.)

compounds containing two or more amino acids linked by the carboxyl group of one amino acid to the amino group of another).

A FRET biosensor based on a QD-protein/peptide conjugate was developed by Medintz et al. (2003) and Clapp et al. (2004, 2005) for detecting maltose (malt sugar, white crystalline sugar, $C_{12}H_{22}O_{11}$, a disaccharide formed from two units of glucose joined with an $\alpha(1\rightarrow4)$ bond). Multiple copies of *Escherichia coli* maltose-binding protein (MBP) (a part of the maltose/maltodextrin system of *E. coli*, responsible for the uptake and efficient catabolism of maltodextrins) coordinated to each QD by a C-terminal oligohistidine segment (a construct in which an *oligohistidine* tag is placed at the *C terminus*) and functioned as sugar receptors. Histidine (abbreviated as His or H), $C_6H_9N_3O_2$ is an essential amino acid in humans; oligonucleotide is a short nucleic acid polymer.

Sensors were self-assembled in solution in a controlled fashion. In one configuration, a β-cyclodextrin-QSY9 dark quencher conjugate bound in the MBP saccharide binding site results in FRET quenching of QD photoluminescence. Cyclodextrins are a family of compounds made up of sugar molecules bound together in a ring (cyclic oligosaccharides), produced from

starch by enzymatic conversion. QSY9 is succinimidyl ester, a fluorescent dye used to stain biological specimens.

Added maltose displaces the β-cyclodextrin-QSY9, and increases the fluorescence of QD in proportion to maltose concentration.

To overcome inherent QD donor–acceptor distance limitations, the sensors were modified to operate on a two-step FRET mechanism by labeling the MBP with another fluorescence dye (Zhou and Ghosh 2006). The organic dye QSY-9 (absorption ~ 560 nm) labeled β-CD (β-cyclodextrin) was used to bind to MBP attached to the QD560 (emission ~ 560 nm) surface such that the fluorescence was quenched, Figure 6.15a. With the addition of free maltose, displacement of the β-CD-QSY9 occurred. As a consequence, there was fluorescence recovery from the QD. The optimized sensor contained 10 copies of β-CD-QSY9 bound to the QD complex. Here, 75% of the QD fluorescence
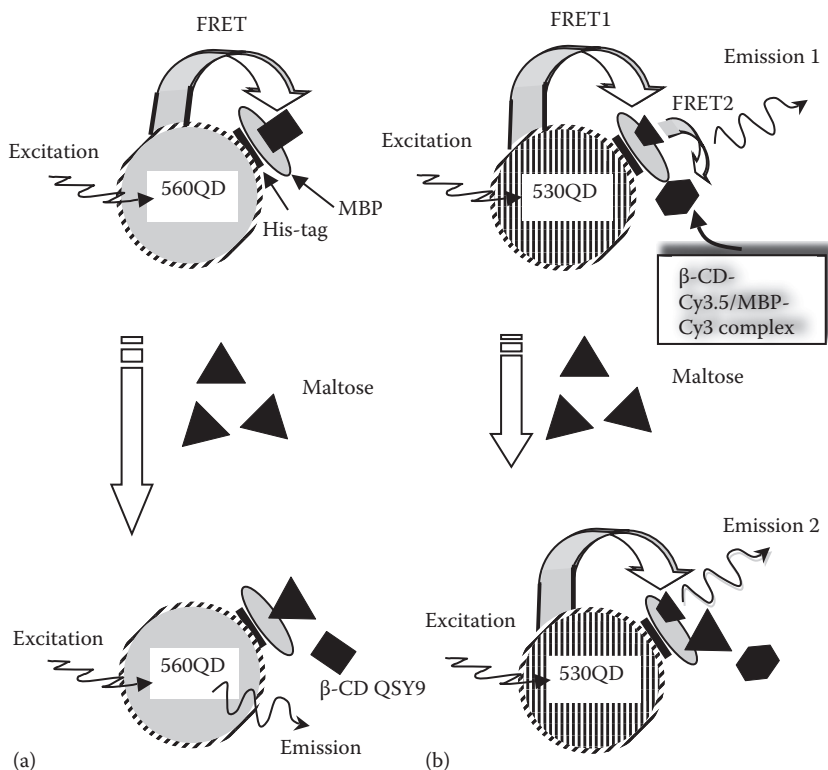
FIGURE 6.15
QD-based maltose nanosensor. (a) *Above:* β-CD-QSY-9/MBP complex is bound to QD through a peptide His-tag. QD fluorescence is quenched by QSY-9. *Below:* The displacement of β-CD-QSY-9 with maltose results in the recovery of QD fluorescence. (b) *Above:* β-CD-Cy3.5/MBP-Cy3 complex is bound to QD through a peptide His-tag. Cy3.5 fluorescence is emitted through a two-step FRET process. *Below:* The displacement of β-CD-Cy3.5 with maltose results in Cy3 fluorescence emission. (After Zhou, M. and Ghosh, I., *Pept. Sci.*, 88, 325, doi: 10.1002/bip, 2006.)

was quenched. Upon complete displacement of β-CD-QSY9 with maltose, an increase in fluorescence of ~ threefold was observed.

These applications are problematical due to the uncertain distance between QD and acceptor. The FRET efficiency may increase through electron transfer between QDs and organic dyes. To surmount these limitations, a maltose sensor was designed in which 10 copies of Cy3 (absorption 556 nm and emission ~ 570 nm) labeled MBP were first incorporated on the QD530 (emission ~ 530 nm) surface, Figure 6.15b. The second step involved binding of the Cy3.5 (absorption ~ 575 nm and emission ~ 595 nm) labeled β-CD. Fluorescence energy was first transferred from the QD530 to MBP-Cy3. Subsequently, the emission energy of Cy3 was transferred to β-CD-Cy3.5. The displacement of β-CD-Cy3.5 from MBP by maltose produced associated fluorescence fall from Cy3.5 and a rise from Cy3. This geometry permitted sufficient energy to transfer to Cy3. However, it minimized direct energy transfer to Cy3.5. The above-mentioned approaches showed how appropriately designed QD complexes with peptide immobilization tags are useful in determining small molecule concentrations in the 100 nM–10 μM range.

6.11.4 Sensor for Determining the Dissociation Constant (K_d) between Rev and RRE

Rev is a vital HIV-1 regulatory protein within the *env* gene (a *gene* that encodes a protein precursor for the envelope proteins, found in the retroviral genome) of HIV-1 RNA genome (the entirety of an organism's hereditary information). The HIV-1 regulatory protein Rev is expressed early in the virus life cycle and thus may be an important target for the immune control of HIV-1-infection and for effective vaccines. Zhang and Johnson (2006) used QDs instead of organic dyes to develop a sensor to determine the dissociation constant (K_d) between Rev and RRE (the *Rev* response element is a region in the RNA molecule of the HIV *env* gene) based on FRET. The QDs (emission ~ 605 nm) were functionalized with streptavidin, a tetrameric protein that binds very tightly to the small molecule, biotin ($C_{10}H_{16}N_2O_3S$). The streptavidin-coated QDs functioned as both a nano-scaffold and a FRET donor in this QD-based NS. Streptavidin was bound to biotin-labeled RRE. Nano-scaffolding is a medical process used to regrow tissue and bone, including limbs and organs.

The 605QD is an excellent energy donor with Cy5 for several reasons: (i) absence of cross talk between the emission spectra of 605QD and that of Cy5; (ii) absence of direct excitation of Cy5 at the wavelength of 488 nm; (iii) 605QD has a high quantum yield (~0.6), Cy5 has a high extinction coefficient (~250,000 M^{-1} cm^{-1}); and (iv) a single 605QD can efficiently couple to multiple Cy5-labeled Rev–RRE complexes around its surface enabling efficient FRET even at distance approaching $2 \times R_0$ for this 605QD/Cy5 FRET pair.

The RRE functionalized QDs effectively complexed multiple copies of a Cy5 labeled 17-mer peptide containing the arginine ($C_6H_{14}N_4O_2$)-rich region

of Rev; arginine is an amino acid, obtained from the hydrolysis or digestion of plant and animal protein.

This reduced the QD fluorescence and increased Cy5 fluorescence. QD fluorescence was recovered by titrating unlabeled Rev peptide into the FRET complex. The K_d of the Rev–RRE complex was estimated by monitoring the fluorescence change of QD and Cy5 in this titration process. Identical results were obtained from the titration curves from the two fluorophores (loss in FRET and increase in non-FRET emission) giving K_d ~ 11–14 nM. This afforded a simple method for monitoring Rev–RRE interaction and possibly any peptide/RNA, peptide/DNA, or peptide–protein interaction.

This QD-based NS offers a convenient approach for sensitive Rev–RRE interaction assay due to its exceptional characteristics of noninterference in the native interaction of Rev with RRE, improved FRET efficiency, high sensitivity, and FRET-related two-parameter detection using the instrumentation of standard fluorescence spectrometer and available streptavidin-coated QDs. It is used to study the effect of inhibitor upon Rev–RRE interaction, and likely to have more applications in the development of new drugs against HIV-1 infection

The application of QDs as FRET acceptors is not suggested because of their broad absorption bands, which favor excitation cross talk. Generally, FRET applications of QDs should only be considered if there is another QD-specific advantage for the system in question, such as the possibility to avoid excitation cross talk, their longer fluorescence lifetimes or their very large two-photon action cross sections. Mostly, fluorescent proteins or organic dyes are to be favored for FRET.

6.12 Electrochemiluminescent Nanosensors for Remote Detection

Electrochemiluminescence, also called electrogenerated chemiluminescence, is a kind of luminescence produced during electrochemical reactions in solutions. It is a process in which highly reactive species are generated from stable precursors at the surface of an electrode by energetic electron transfer (redox) reactions of electrogenerated species. These highly reactive species react with one another producing light.

Electrochemiluminescent (ECL) is usually observed during the application of potential (several volts) to electrodes of electrochemical cell that contains solution of luminescent species. Combining analytical advantages of chemiluminescent analysis (absence of background optical signal) with ease of reaction control by applying electrode potential, ECL detection technology promises scientists new "yardsticks" for quantification.

Chovin et al. (2006) reported a micro sensing device for ECL imaging through an ordered microarray of optical apertures. This microdevice,

fabricated by wet chemical etching of an optical fiber bundle comprising 6000 individually clad 3–4 μm diameter optical fibers, consisted of a systematically planned array of etched optical fiber cores, each tip of which was surrounded by a ring-shaped gold nanoelectrode (Figure 6.16). By this method of fabrication, the initial architecture of the optical fiber bundle was retained and thus the microarray maintained its imaging properties. An electrophoretic paint was deposited cathodically onto the surface of the nanotip array. The electrophoretic paint is a coating formed on a cathode as a result of electrophoresis (the migration of charged colloidal particles or molecules through a solution under the influence of an applied electric field) and coagulation of colloidal particles. By carefully controlling the deposition time of the paint, the array
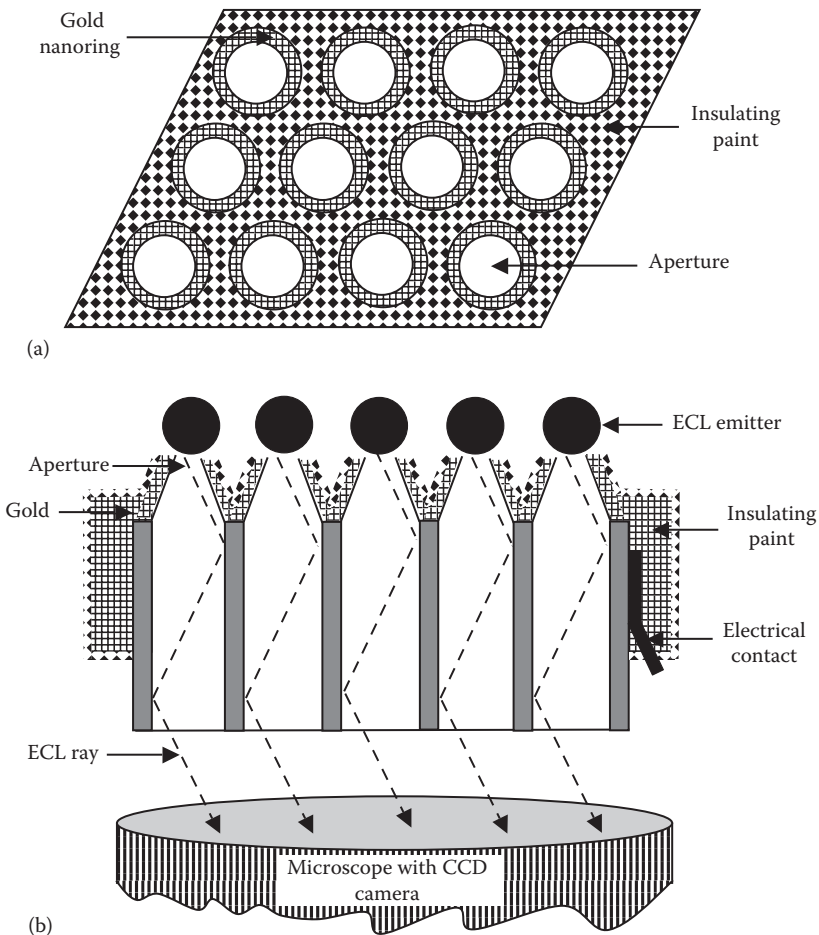
FIGURE 6.16

Microarray of electroluminescent nanosensors: (a) top view and (b) side view. (After Chovin, A. et al., *Meas. Sci. Technol.*, 17, 1211, 2006.)

was covered with insulating paint. Heating of the resin-coated array resulted in a shrinking of the resin (clear to translucent, solid, or semisolid viscous substance) film and exposure of the metal tips of the fiber bundle. To permit light transmission through the etched fiber cores, the exposed gold layer was removed by reacting in a gold-etching solution. The gold film accomplished two functions: It served to confine light to the tip apex and also as the electrode material to perform electrochemical reactions. The high-density microarray comprised 6000 NSs. Each sensor was a sub-wavelength aperture formed at the apex of an etched fiber core and surrounded by a gold-nanoring electrode. Such nanostructured microdevice integrated ECL-light generation, collection, and imaging in a microarray format.

The microdevice was applied to the ECL detection of NADH. The NADH is an abbreviation for the reduced form of NAD (nicotinamide adenine dinucleotide), a coenzyme found in all living cells. Direct imaging of NADH was performed through the NS microarray itself. The ECL microarray showed good temporal stability and reproducibility. The substrate concentration was monitored and imaged by measuring NADH ECL intensity. Therefore, many clinically important analytes are detectable by changing the nature of the dehydrogenase enzyme.

Chemiluminescence-based sensor for hydrogen peroxide was made of peroxalate ester NPs and encapsulated fluorescent dyes (Lee et al. 2007). Hydrogen peroxide has a reputation as a topical germ killer, but it is gaining fame in the medical community as an early indicator of disease in the body. It is an important molecule involved in physiological signaling process. Hydrogen peroxide is thought to be overproduced by cells at the early stages of most diseases. Overproduction of hydrogen peroxide causes tissue damage and inflammation.

The sensing mechanism is as follows: When the nano particles bump into hydrogen peroxide, the latter reacts with the peroxalate ester groups, producing a high-energy dioxetanedione intermediate. The chemical compound 1,2-dioxetanedione, often called peroxyacid ester, is an unstable oxide of carbon (an oxocarbon) with formula C_2O_4. It chemically stimulates encapsulated fluorescent dyes, such as pentacene, exciting chemiluminescence. The emitted photons (or light) can be detected. The chemiluminescence wavelength could be tuned by changing the encapsulated fluorescent dyes.

These NPs are incredibly sensitive, so one can detect nanomolar concentrations of hydrogen peroxide. Such NPs could someday be used as simple, all-purpose diagnostic tools for most diseases, for example, the NPs would be injected by needle into a certain area of the body such as the heart. If the nanoparticles encountered hydrogen peroxide, they would emit light. When a doctor notices a significant amount of light activity in the region, the doctor would know that the patient may be presenting early signs of a disease in that area of the body. Further, these NPs penetrate deep into the tissue and operate at a high wavelength, making them sensitive indicators of the presence of hydrogen peroxide produced by any kind of inflammation.

6.13 Crossed Zinc Oxide Nanorods as Resistive UV Nanosensors

UV photosensors find wide-ranging applications in medicine, ecology, space communications, high-temperature plasma research, chemical and biological sensing, and in military for flame and missile launching detection. UV detectors are used to monitor and determine the earth's ozone layer thickness. Although several kinds of photosensors are available now for the UV range, the development of NSs for these applications will enable more portable detectors. Usually, direct wide-band-gap materials like GaN, ZnSe, and AlGaN are used as sensing materials. Analogous to III–V materials, ZnO also possesses high UV photosensitivity, which is most important for UV photodetection. Thus, zinc oxide is also a potentially strong UV-sensing material.

Chai et al. (2009) reported a self-assembled ZnO nanocross–based UV photosensor (Figure 6.14). They showed a simple and cost-effective method to fabricate a crossed ZnO nanorods device, possibly leading to a next generation of photosensors for a wide range of applications. The crossed ZnO nanorods (Figure 6.17) were synthesized in a hydrothermal reactor with an aqueous solution deposition technique. "Hydros" meaning water and "thermos" meaning heat, hydrothermal processing is defined, in general, as any heterogeneous reaction in the presence of aqueous/nonaqueous solvents or mineralizers under high pressure and temperature conditions to dissolve and recrystallize (recover) materials that are relatively insoluble under ordinary conditions. The hydrothermal technique is an important tool for advanced materials processing, particularly owing to its advantages in the processing of nanostructural materials for a wide variety of technological applications
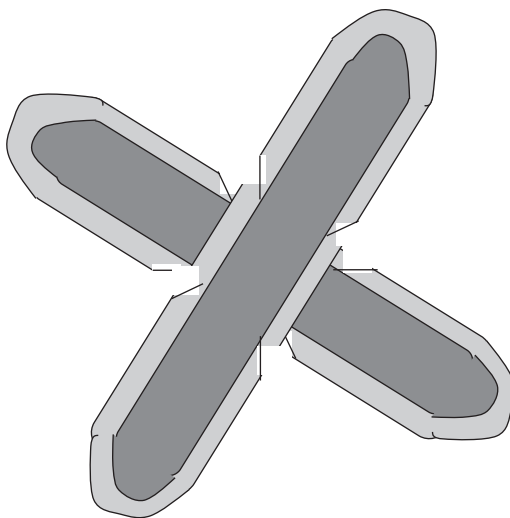
FIGURE 6.17
Crossed ZnO nanorod. (After Chai, G. et al., *Sens. Actuators, A*, 150, 184, 2009.)

such as electronics, optoelectronics, catalysis, ceramics, magnetic data storage, biomedical, and biophotonics. Zinc sulfate $Zn(SO_4) \cdot 7H_2O$ and ammonia solution NH_4OH (29.6%) were used as starting materials. Transmission electron microscopy (TEM) confirmed that the nanorods were crystalline. An *in situ* lift-out technique was used to fabricate this photosensor based on self-assembled crossed nanorods. A focused ion beam system was used to contact single crossed ZnO nanorods grown by an aqueous solution process. In order to realize the easy attachment, the nanorod was deposited and micro/nano manipulated on to a SiO_2/Si substrate already processed to have respective distribution of pre-patterned external electrodes. The UV sensitivity was measured using a two-terminal ZnO nanorod device. The current–voltage (*I*–*V*) characteristics showed linear behavior. The photosensor exhibited a response of ~15 mA W^{-1} for UV light (361 nm) under 1 V bias. Response measurements showed that such a photosensor was suitable for low levels of UV detection. Widening the band gap by doping with Mg is a possible future research direction in order to cover UV-A (315–400 nm), UV-B (280–315 nm), and UV-C (100–280 nm) regions of the UV spectrum. The prototype device provided a simple method for nanowire synthesis and demonstrates the possibility of constructing nanoscale photodetectors for nano-optics applications.

6.14 Discussion and Conclusions

Important information regarding nanomaterials exploited, principles of operation, and applications of various optical NSs described in this chapter is summarized in Table 6.1. Inspection of Table 6.1 shows that a plurality of approaches has been followed for the development of optical NSs. Metallic NPs like gold, silver, and copper constitute interesting sensing tools by providing a physical output such as a change in color from a chemical input, for example, the chemically induced modification of properties of the particle. The position of the surface plasmon band of metallic NPs in solution, that is, the color of the solution provides a vital clue but enhancement of the signal by Raman spectroscopy has given birth to a new field known as SERS. The introduction of SERS has enabled single-molecule detection. Highly portable, user-friendly instruments are available. A simplified replacement of SPR instrument was reported in which an optically transparent substrate was used to enable transmission mode operation. Colloidal gold NPs were immobilized on the substrate. Color changes were produced by the sample. The device was coupled to a UV–visible spectrometer.

Building noninvasive or minimally invasive NSs is the goal that all nanotechnologists will strive to reach. In all these sensors, nano dimension is the main theme. Fiber optic NSs for living cell examination are an attempt to visualize the cellular phenomena in living condition. Sensors for VOCs employ

TABLE 6.1

Optical Nanosensor Family

Sl. No.	Name of the Nanosensor	Nanomaterial Used	Operating Principle	Sensed Quantity/ Applications
1.	Localized surface plasmon resonance (LSPR) spectroscopy	Noble-metal nanoparticles	Refractive index–based detection: wavelength shift in the LSPR extinction or scattering maximum	Chemical and biological targets
2.	Surface-enhanced Raman scattering (SERS)	Metallic nanostructures	LSPR, fluorescence, and Raman spectroscopy	Biosensing applications
3.	Colorimetric gold nanoparticle spectrophotometric sensor	Colloidal gold on an optically transparent substrate	A colorimetric signal due to the changes in surface plasmon absorbance.	Label-free detection of biomolecular interactions
4.	Fiber-optic nanosensor	Optical fibers and SWCNTs; in-fiber nanocavity doped with vapochromic material	Absorbance, reflectance, fluorescence, surface plasmon resonance (SPR), refractive index and colorimetry	Chemical and biological applications
5.	Nanograting-based optical accelerometer	Acceleration	Nanogratings	Modulation of the effective refractive index of the structures
6.	Fluorescent pH-sensitive nanosensor	A pH sensitive fluorescent dye along with a pH insensitive reference dye embedded in a polymer matrix	Responses from the pH sensitive dye and the reference dye provide pH measurements by fluorescence ratio imaging microscopy	Physiological levels of pH and sodium
7.	PEBBLE nanosensors	Fluorescent indicators encased in a permeable biologically inert matrix	Ratiometric analysis	Bioanalysis
8.	Quantum dots as fluorescent labels	QD	Labeling	Biological
9.	Quantum dot FRET-based probes	QD	FRET	Biological
10.	ECL nanosensors	Microarray of optical apertures	Electrochemiluminescence	Chemical and biological
11.	ZnO nanorod ultraviolet nanosensor	ZnO nanorods	Current per watt change under voltage bias	UV photosensing

carbon nanotubes as high sensitivity nanomaterials with optical fiber technology to build Fabry–Perot interferometers. A nanopipette giving ratiometric or intensity-based responses to pH and sodium was reported. A polymeric fluorescent pH-NS compares the signals from a pH-sensitive dye with that from a reference dye to obtain pH of the analyte. Both are housed in a polymeric casing, hence the name. It is useful for physiological pH monitoring. PEBBLE NSs are also aimed at obtaining responses from living beings without disturbing the normal functioning. QDs are used as substitutes for organic fluorophores in a variety of applications. ECL NS arrays are useful for imaging purposes. Thus, in all the sensors described in this chapter, either special nanoscale optical phenomena are exploited for sensitivity improvement or sensor size is downscaled to avoid disturbance to the sensed environment.

Review Exercises

6.1 An ideal bioanalytical and biomedical sensor should achieve, in live cells and *in vivo*, real-time tracking of biological/chemical/physical processes as well as detection of disease-related abnormal features, with no interferences. Are traditional sensors like microelectrodes or fiber-optical sensors close to the ideal sensor? Argue.

6.2 Mention three intracellular delivery methods of nanoparticles through the plasma membrane barrier. Can *in vivo* delivery of the nanoparticles to cells be done by intravenous injection?

6.3 Some types of nanoparticles possess unique but controllable optical/ magnetic properties that are superior to molecular probes. Give two examples.

6.4 Can the LSPR wavelength of the metallic nanoparticles be tuned by changing the shape, size, and composition of the metal nanoparticle or metal shell thickness?

6.5 Do the optical or magnetic characteristics of QDs and nanoparticles themselves change in response to specific analytes? What is the effect of labeling with suitable dyes, ligands or receptors?

6.6 FRET-based maltose sensors were developed using QD, MBP and a dye complex that forms a FRET pair with QD. The dye complex, β-cyclodextrin-acceptor dye conjugates, is initially bound to MBP, resulting in FRET quenching of the QD fluorescence. Added maltose displaces the dye complex. What is the influence of maltose concentration on the fluorescence of QD? Explain.

6.7 An important phenomenon is defined as follows: A distance-dependent energy transfer mechanism between two dye molecules by a

nonradiative, long-range dipole–dipole coupling mechanism. Name the phenomenon.

6.8   In a measurement technique, a signal is measured with respect to a reference signal. What is it called?

6.9   What is the full form of: (i) LSPR, (ii) SPR, (iii) SERS, (iv) PEBBLE, and (v) SPIO?

6.10 Give reasons why ZnO has attracted great interest as a UV photosensor material?

6.11 Point out the advantages of PEBBLE scheme over other means of cellular measurement with reference to the size, spatial resolution, interference from background fluorescence, protection of the recognition elements and cell, and co-location of multiple fluorophores.

6.12 Mention some methods for delivery of PEBBLEs into cells. Does the polymer matrix of PEBBLEs allow macromolecules such as protein to diffuse across it?

6.13 Is SERS a vibrational spectroscopic method or a refractive index–based detection scheme? Does the SERS intensity depend upon the excitation of the LSPR?

6.14 The Langmuir–Blodgett technique is very useful to deposit ultrathin organic films with a precise control over the architecture of the films at the molecular level. In light of this, discuss its application to optical fiber sensor for SWCNT deposition on the distal face. Draw and explain the optical fiber interrogation system.

6.15 In an optical fiber nanobiosensor, the optical fiber's tip is significantly less than the wavelength of light used to excite the target analyte. So, the photons cannot escape from the tip of the fiber by conventional optics. After the photons have traveled as far down the fiber as possible, how is the excitation for the fluorescent species of interest provided in the vicinity of the biosensing layer? Discuss the role of evanescent fields here.

References

Aylott, J. W. 2003. Optical nanosensors—An enabling technology for intracellular measurements. *Analyst* 128: 309–312, doi: 10.1039/b302174m

Barbillon, G., J.-L. Bijeon, J. Plain, M. L. de la Chapelle, P.-M. Adam, and P. Royer. 2007. Biological and chemical gold nanosensors based on localized surface plasmon resonance. *Gold Bulletin* 40(3): 240–244.

Borisov, S. M. and I. Klimant. 2008. Optical nanosensors—Smart tools in bioanalytics. *Analyst* 133: 1302–1307, doi: 10.1039/b805432k

Cao, Y., Y. E. L. Koo, S. M. Koo, and R. Kopelman. 2005. Ratiometric singlet oxygen nano-optodes and their use for monitoring photodynamic therapy nanoplat-forms. *Photochemistry and Photobiology* 81(6): 1489–1498 [PubMed: 16107183].

Cao, Y., Y. E. L. Koo, and R. Kopelman. 2004. Poly(decyl methacrylate)-based fluorescent PEBBLE swarm nanosensors for measuring dissolved oxygen in biosamples. *Analyst* 129: 745–750 [PubMed: 15284919].

Carr, D. W., G. R. Bogart, and B. E. N. Keeler. 2004. Femto-photonics: Optical transducers utilizing novel sub-wavelength dual layer grating structure. In: *Proceedings of the Hilton Head Solid-State Sensors, Actuators, and Microsystems Workshop*, Hilton Head Island, SC, June 2004, pp. 91–92.

Chai, G., O. Lupan, L. Chow, and H. Heinrich. 2009. Crossed zinc oxide nanorods for ultraviolet radiation detection. *Sensors and Actuators A* 150: 184–187.

Cheng, Z. L. and C. A. Aspinwall. 2006. Nanometre-sized molecular oxygen sensors prepared from polymer stabilized phospholipid vesicles. *Analyst* 131(2): 236–243 [PubMed: 16440088].

Chovin, A., P. Garrigue, G. Pecastaings, H. Saadaoui, and N. Sojic. 2006. Development of an ordered microarray of electrochemiluminescent nanosensors. *Measurement Science and Technology* 17: 1211–1219.

Clapp, A. R., I. L. Medintz, J. M. Mauro, B. R. Fisher, M. G. Bawendi, and H. Mattoussi. 2004. Fluorescence resonance energy transfer between quantum dot donors and dye-labeled protein acceptors. *Journal of the American Chemical Society* 126(1): 301–310.

Clapp, A. R., I. L. Medintz, H. T. Uyeda, B. R. Fisher, E. R. Goldman, M. G. Bawendi, and H. Mattoussi. 2005. Quantum dot-based multiplexed fluorescence resonance energy transfer. *Journal of the American Chemical Society* 127: 18212–18221.

Clark, H. A., M. Hoyer, M. A. Philbert, and R. Kopelman. 1999. Optical nanosensors for chemical analysis inside single living cells. 1. Fabrication, characterization, and methods for intracellular delivery of PEBBLE sensors. *Analytical Chemistry* 71: 4831–4836.

Consales, M., S. Campopianoa, A. Cutoloa, M. Penza, P. Aversab, G. Cassanob, M. Giordanoc, and A. Cusano. 2006. Carbon nanotubes thin films fiber optic and acoustic VOCs sensors: Performances analysis. *Sensors and Actuators B* 118: 232–242.

Consales, M., A. Cutolo, M. Penza, P. Aversa, G. Cassano, M. Giordano, and A. Cusano. 2007. Carbon nanotubes coated acoustic and optical VOCs sensors: Towards the tailoring of the sensing performances. *IEEE Transactions on Nanotechnology* 6(6): 601–612.

Consales, M., A. Cutolo, M. Penza, P. Aversa, M. Giordano, and A. Cusano. 2008. Fiber optic chemical nanosensors based on engineered single-walled carbon nanotubes. *Journal of Sensors*, 2008: Article ID936074, 29, doi: 10.1155/2008/936074

Coupland, P. G., S. J. Briddon, and J. W. Aylott. 2009. Using fluorescent pH-sensitive nanosensors to report their intracellular location after Tat-mediated delivery. *Integrative Biology* 1: 318–323.

Cusano, A., M. Giordano, A. Cutolo, M. Pisco1, and M. Consales. 2008. Integrated development of chemoptical fiber nanosensors. *Current Analytical Chemistry* 4: 296–315.

Dubertret, B., P. Skourides, D. J. Norris, V. Noireaux, A. H. Brivanlou, and A. Libchaber. 2002. *In vivo* imaging of quantum dots encapsulated in phospholipid micelles. *Science* 298: 1759–1762.

Elosua, C., I. R. Matias, C. Bariain, and F. J. Arregui. 2006. Development of an in-fiber nanocavity towards detection of volatile organic gases. *Sensors* 6: 578–592.

Guice, K. B., M. E. Caldorera, and M. J. McShane. 2005. Nanoscale internally referenced oxygen sensors produced from self-assembled nanofilms on fluorescent nanoparticles. *Journal of Biomedical Optics* 10: 64031-1 to 64031-10.

Hohng, S. and T. Ha. 2005. Single-molecule quantum-dot fluorescence resonance energy transfer. *Chemphyschem* 6: 956–960.

Jackson, J. B. and N. J. Halas. 2004. Surface-enhanced Raman scattering on tunable plasmonic nanoparticle substrates. *PNAS* 101(52): 17930–17935.

Jamieson, T., R. Bakhshi, D. Petrova, R. Pocock, M. Imani, and A. M. Seifalian. 2007. Biological applications of quantum dots. *Biomaterials* 28: 4717–4732.

Javier, A. M., P. D. Pino, S. Kudera, and W. J. Parak. 2009. In *Biosensing Using Nanomaterials,* Merkoci, A. (Ed.), John Wiley & Sons, New Jersey, pp. 256–270.

King, M. and R. Kopelman. 2003. Development of a hydroxyl radical ratiometric nanoprobe. *Sensors and Actuators B* 90: 76–81.

Kneipp, J., H. Kneipp, B. Wittig, and K. Kneipp. 2010a. Novel optical nanosensors for probing and imaging live cells. *Nanomedicine: Nanotechnology, Biology, and Medicine* 6: 214–226.

Kneipp, J., B. Wittig, H. Bohr, and K. Kneipp. 2010b. Surface-enhanced Raman scattering: A new optical probe in molecular biophysics and biomedicine. *Theoretical Chemistry Accounts* 125: 319–327.

Koo, Y., Y. Cao, R. Kopelman, S. M. Koo, M. Brasuel, and M. A. Philbert. 2004. Real-time measurements of dissolved oxygen inside live cells by ormosil (organically modified silicate) fluorescent pebble nanosensors. *Analytical Chemistry* 76: 2498–2505 [PubMed: 15117189].

Kreft, O., A. M. Javier, G. B. Sukhorukovac, and W. J. Parak. 2007. Polymer microcapsules as mobile local pH-sensors. *Journal of Materials Chemistry* 17: 4471–4476.

Krishnamoorthy, U., R. H. Olsson III, G. R. Bogart, M. S. Baker, D. W. Carr, T. P. Swiler, and P. J. Clews. 2008. In-plane MEMS-based nano-g accelerometer with sub-wavelength optical resonant sensor. *Sensors and Actuators A: Physical* 145–146: 283–290.

Lee, D., S. Khaja, J. C. Velasquez-Castano, M. Dasari, C. Sun, J. Petros, W. R. Taylor, and N. Murthy. 2007. *In vivo* imaging of hydrogen peroxide with chemiluminiscent nanoparticles. *Nature Materials* 6(10): 765–769 [PubMed: 17704780].

Lee, Y.-E. K., R. Kopelman, and R. Smith. 2009. Nanoparticle PEBBLE sensors in live cells and in vivo. *Annu. Rev. Anal. Chem.* 2: 57–76, doi: 10.1146/annurev.anchem.1.031207.112823.

Martin-Orozco, N., N. Touret, M. L. Zaharik, E. Park, R. Kopelman, S. Miller, B. B. Finlay, P. Gros and S. Grinstein. 2006. Visualization of vacuolar acidification-induced transcription of genes of pathogens inside macrophages. *Molecular Biology of the Cell* 17(1): 498–510 [PubMed: 16251362].

McKinney, S. A., A. C. Declais, D. M. J. Lilley, and T. Ha. 2003. Structural dynamics of individual holliday junctions. *Nature Structural Biology* 10: 93–97.

Medintz, I. L., A. R. Clapp, H. Mattoussi, E. R. Goldman, B. Fisher, and J. M. Mauro. 2003. Self-assembled nanoscale biosensors based on quantum dot FRET donors. *Nature Materials* 2: 630–638.

Nath, N. and A. Chilkoti. 2002. A colorimetric gold nanoparticle sensor to interrogate biomolecular interactions in real time on a surface. *Analytical Chemistry* 74: 504–509.

Netti, C. and H. Stanford. 2006. Applications of reproducible SERS substrates for trace level detection. *Spectroscopy* Jun 1, 2006 pp. 8–17.

Pandana, H., K. H. Aschenbach, and R. D. Gomez. 2008. Systematic aptamer-gold nanoparticle colorimetry for protein detection: Thrombin. *IEEE Sensors Journal* 8(6): 661–666.

Penza, M., G. Cassano1, P. Aversa, A. Cusano, A. Cutolo, M. Giordano, and L. Nicolais. 2005. Carbon nanotube acoustic and optical sensors for volatile organic compound detection. *Nanotechnology* 16: 2536–2547.

Piper, J. D., R. W. Clarke, Y. E. Korchev, L. Ying, and D. Klenerman. 2006. A renewable nanosensor based on a glass nanopipette. *Journal of the American Chemical Society* 128: 16462–16463.

Reddy, G. R., M. S. Bhojani, P. McConville, J. Moody, B. A. Moffat, D. E. Hall, G. Kim, Y. E. Koo, M. J. Woolliscroft, J. V. Sugai, T. D. Johnson, M. A. Philbert, R. Kopelman, A. Rchemtulla, and B. D. Ross. 2006. Vascular targeted nanoparticles for imaging and treatment of brain tumors. *Clinical Cancer Research* 12(22): 6677–6686 [PubMed: 17121886].

Resch-Genger, U., M. Grabolle, S. Cavaliere-Jaricot, R. Nitschke, and T. Nann. 2008. Quantum dots versus organic dyes as fluorescent labels. *Nature Methods* 5(9): 763–775.

Shi, L., V. De Paoli, N. Rosenzweig, and Z. Rosenzweig. 2006. Synthesis and application of quantum dot FRET-based protease sensors. *Journal of the American Chemical Society* 128: 10378–10379.

Sumner, J. P., N. M. Westerberg, A. K. Stoddard, C. A. Fierke, and R. Kopelman. 2006. Cu^+- and Cu^{2+}-sensitive PEBBLE fluorescent nanosensors using DsRed as the recognition element. *Sensors and Actuators B* 113: 760–767.

Sun, H., A. M. Scharff-Poulsen, H. Gu, and K. Almdal. 2006. Synthesis and characterization of ratiometric, pH sensing nanoparticles with covalently attached fluorescent dyes. *Chemistry of Materials* 18(15): 3381–3384.

Talley, C. E., L. Jusinski, C. W. Hollars, S. M. Lane, and T. Huser. 2004. Intracellular pH sensors based on surface enhanced Raman scattering. *Analytical Chemistry* 76: 7064–7068 [PubMed: 15571360].

Vo-Dinh, T. and P. Kasili. 2005. Fiber-optic nanosensors for single-cell monitoring. *Analytical and Bioanalytical Chemistry* 382: 918–925.

Webster, A., P. Coupland, F. D. Houghton, H. J. Leese, and J. W. Aylott. 2007. The delivery of PEBBLE nanosensors to measure the intracellular environment. *Biochemical Society Transactions* 35(3): 538–543.

Yonzon, C. R., X. Zhang, J. Zhao, and R. P. Van Duyne. 2007. Surface-enhanced nanosensors. *Spectroscopy* 22(1): 42–56.

Zhang, C.-Y. and L. W. Johnson. 2006. Quantum-dot-based nanosensor for RRE IIB RNA-Rev peptide interaction assay. *Journal of the American Chemical Society* 128(16): 5324–5325, doi: 10.1021/ja060537y

Zhou, M. and I. Ghosh. 2006. Quantum dots and peptides: A bright future together. *Peptide Science* 88(3): 325–339, doi: 10.1002/bip

Zwiller, V., N. Akopian, M. van Weert, M. van Kouwen, U. Perinetti, L. Kouwenhoven, R. Algra, J. Gómez Rivas, E. Bakkers, G. Patriarche, L. Liu, J.-C. Harmand, Y. Kobayashi, and J. Motohisa. 2008. Optics with single nanowires. *C. R. Physique* 9: 804–815.

7

Magnetic Nanosensors

7.1 Introduction

What is a magnetic sensor? It is a transducer that converts a magnetic field into an electrical signal.

A magnetic sensor is also required whenever a nonmagnetic signal is detected by means of an intermediary signal detection into the magnetic signal domain in a tandem transducer, for example, the detection of a current through its magnetic field or the mechanical displacement of a magnet. Therefore, two groups of *direct and indirect magnetic sensors* are distinguishable.

Where are magnetic sensors used? Magnetic sensors have been employed for a long time for applications such as current sensing, encoders, gear tooth sensing, linear and rotary position sensing, rotational speed sensing, motion sensing, and so on. Particularly, solid-state magnetic field sensors, which are generally utilized for this purpose, have an inherent advantage in size and power consumption compared with search coil, fluxgate, and more intricate low-field sensing techniques, such as superconducting quantum interference detectors (SQUIDs) and spin resonance magnetometers. This holds true even for high-frequency applications.

The SQUID is an extremely sensitive magnetic flux-to-voltage transducer. The main component of the SQUID is the Josephson junction, comprising essentially two superconductors weakly coupled through a small insulating gap or constriction. This junction has unique electrical/magnetic properties. The SQUID device consists of two superconductors separated by thin insulating layers to form two parallel Josephson junctions. Two such superconductors separated by a thin insulating layer experience tunneling of Cooper pairs of electrons (two electrons that are bound together at low temperatures with correlated motion and constituting a single entity) through the junction. The Cooper pairs on each side of the junction are represented by a wave function similar to a free particle wave function. In the DC Josephson effect, a current proportional to the phase difference of the wave functions flows in the junction in the absence of a voltage.

The SQUID device may be configured as a magnetometer (a scientific instrument used to measure the strength and/or direction of the magnetic field) to detect incredibly small magnetic fields. A superconducting pick-up coil couples the SQUID to the ambient magnetic field. If a constant biasing current is maintained in the SQUID device, the measured voltage oscillates

with the changes in phase at the two junctions, which depends upon the change in the magnetic flux.

An electron spin resonance magnetometer utilizes the electron paramagnetic resonance for measuring the instantaneous value of the time-varying magnetic field.

Returning to our discussion on solid-state magnetic sensors, *what is the working principle of solid-state magnetic sensors?* These sensors work on the principle of converting the magnetic field into a voltage or resistance using a DC supply. The sensing can be done in an extremely small, lithographically patterned area, with additional reduction of size and power requirements. The diminutive size of a solid-state element increases the resolution for fields that vary over small distances. It also allows the enclosure of arrays of sensors in a small package, further decreasing size and power requirements.

What has been the industrial impact of solid-state magnetic sensors? Solid-state magnetic field sensors have revolutionized measurement and control systems. Industry continues to reap the benefits of solid-state magnetic field sensing. Every day new applications are found for solid-state magnetic field sensors.

One of the key types of solid-state magnetic field sensors is the magnetoresistive sensor.

7.2 Magnetoresistance Sensors

What is magnetoresistance (*MR*)? Magnetoresistance is the property of a material or system of materials that results in a change of resistance when they are exposed to a magnetic field (Freitas et al. 2007, Koh and Josephson 2009). *Is magnetoimpedance, a phenomenon involving the change of total impedance of a ferromagnetic conductor in a magnetic field, when a high frequency alternating current flows through it, considered as magnetoresistance?* No.

The discovery of large magnetoresistive effects has led to the development of solid-state magnetic sensors that can replace high-priced wire-wound sensors in various applications. The usual figure of merit for magnetoresistance is the *magnetoresistance (MR) ratio* traditionally defined in terms of maximum resistance R_{max} and minimum resistance R_{min}, by (Jander et al. 2005)

$$\text{MR\%} = \left\{ \frac{R_{max} - R_{min}}{R_{min}} \right\} \times 100 \qquad (7.1)$$

By this definition, can MR ratio exceed 100%? Yes.

What does the MR ratio indicate? The MR ratio indicates the maximum signal that can be obtained from the sensor.

Example 7.1

In a magnetoresistor, the resistance changed from 700 to 710 Ω. Find the MR ratio.

From Equation 7.1,

$$\text{MR\%} = \left\{ \frac{R_{max} - R_{min}}{R_{min}} \right\} \times 100 = \left\{ \frac{710 - 700}{700} \right\} \times 100 = 1.43\% \qquad (7.2)$$

7.2.1 Ordinary Magnetoresistance: The Hall Effect

All conductors exhibit a weak MR effect known as *ordinary magnetoresistance* (OMR). One such effect is the *Hall effect*: the generation of a voltage in a current-carrying conductor placed in a magnetic field; the voltage produced is perpendicular to both the electric current and magnetic field. Its origin lies in the Lorentz force. *What is Lorentz force*? The Lorentz force is the force acting on an electrically charged particle moving through a magnetic plus an electric field. It has two vector components, one proportional to the magnetic field and the other proportional to the electric field, to be added vectorially to obtain the total force: (i) The strength of the *magnetic component* is proportional to the charge q of the particle, the speed v of the particle, the intensity B of the magnetic induction, and the sine of the angle between the vectors **v** and **B**. The *direction* of the magnetic component is decided by the right-hand rule: put the right hand along **v** with fingers pointing in the direction of **v** and the open palm toward the vector **B**. Then stretching the thumb of right hand, the Lorentz force acts along it, pointing from your wrist to the tip of your thumb. (ii) The *electric component* of the Lorentz force $= q \cdot \mathbf{E}$ (charge of the particle multiplied by the electric field).

The explanation of Hall effect is as follows: disturbance of the current distribution in the conductor by the Lorentz force, due to the applied magnetic field, acting on the electrons flowing in the conductor, is responsible for the creation of potential difference (voltage) across it. The deviation of the current path due to the magnetic field produces an increase of the current path length leading to an increase of the effective resistance.

7.2.2 Anisotropic Magnetoresistance

Ferromagnetic materials exhibit a larger magnetoresistive effect known as *anisotropic magnetoresistance (AMR)*. When a current is passed through a magnetic conductor, resistance changes occur based on the relative angle between the current and the conductor's magnetization. *Is it displayed by all metallic magnetic materials?* Yes.

Why is it called "anisotropic?" It is termed *anisotropic* because, in contrast to the previously known OMR, it depends on the angle between the electric current and the magnetization direction. Resistance increases when current

is perpendicular to magnetization and decreases when current is parallel to magnetization.

The AMR effect is described as a change in the scattering cross section of atomic orbitals distorted by the magnetic field. The microscopic theory of this dependence is founded on the larger probability of s–d scattering for electrons moving parallel to the magnetization (Andreev and Dimitrova 2005).

Devices use permalloy (nickel–iron magnetic alloy, generically, with about 20% iron and 80% nickel content) as sensing material deposited on Si substrates in a Wheatstone bridge configuration. *How do AMR devices compare with OMR devices?* AMR devices typically have MR ratios of 1%–2% while that of OMR devices is very feeble.

7.2.3 Giant Magnetoresistance

Like other magnetoresistive effects, giant magnetoresistance (GMR) represents the change in electrical resistance of some materials in response to an applied magnetic field. *What is the usual construction of a GMR device?* A GMR device is a sandwich structure consisting of two or more layers of ferromagnetic metal (typically Fe, Co, NiFe, CoFe or related transition metal alloy) separated by ultrathin, nonmagnetic metal spacer layers (Cr, Cu, Au, or Ru); these spacer layers have to be only a few nanometers in thickness and their interfaces must be of the best quality.

It was discovered that upon application of a magnetic field to magnetic metallic multilayers such as Fe/Cr and Co/Cu, in which ferromagnetic layers were separated by nonmagnetic spacer layers, a few nanometers thick, the electric current experienced a strong influence of the relative orientation of the magnetizations of the magnetic layers, resulting in a significant reduction of the electrical resistance of the multilayer (about 50% at 4.2 K). This effect was found to be much larger than other magnetoresistive effects that had ever been observed in metals and was, therefore, called "giant magnetoresistance." The term giant magnetoresistance was coined because the tremendous change in resistance found in GMR far exceeded that of any AMR devices. In Fe/Cr and Co/Cu multilayers, the magnitude of GMR can be higher than 100% at low temperatures. In modern GMR structures, MR levels at room temperature above 200% are achieved (Reig et al. 2009).

In what type of structures is GMR effect noticed? GMR response is observed in magnetic multilayer structures comprising at least a trilayer arrangement with two magnetic layers separated by a spacer layer. *Is it found in structures containing two spacer layers and one magnetic layer?* No.

What are the two basic configurations of GMR effect? Two variants of GMR effect have been applied in devices: Current-in-plane (CIP) and current-perpendicular-to-plane (CPP). GMR devices are typically operated with the sense current in the plane of the films using electrical contacts at the ends of long, often serpentine, lines. Although the magnetoresistance is reduced because of current shunting through the layers, the alternative CPP

configuration has a resistance that is too low for practical circuit applications (Jander et al. 2005).

Figure 7.1a and b shows the structures of the sensor and magnetization directions in the absence of the magnetic field and after applying magnetic field, respectively. In the former case, the magnetizations of the
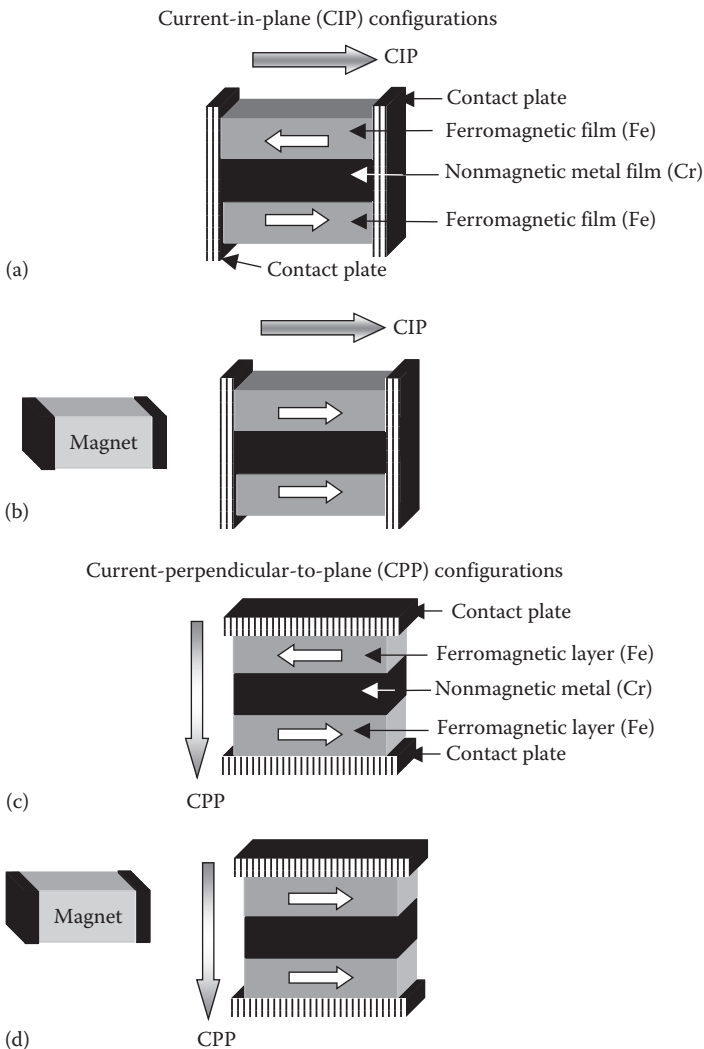
FIGURE 7.1
(a) and (b) Effect of magnetic field on GMR device in current-in-plane (CIP) configuration: (a) magnetizations of ferromagnetic layers are antiparallel in absence of applied magnetic field and (b) the magnetizations are parallel under an applied field. (c) and (d) Effect of magnetic field on GMR device in current-perpendicular-to-plane (CPP) configuration: (c) antiparallel magnetization and (d) parallel magnetization.

ferromagnetic layers are antiparallel (Figure 7.1a), while in the latter case, they become parallel (Figure 7.1b). Here, the current flows parallel to the planes of the layers. Figure 7.1c and d represents the situation for current flowing perpendicular to the planes of layers.

The change in the resistance of the multilayer arises when the applied field aligns the magnetic moments of the successive ferromagnetic layers, as is illustrated schematically in Figure 7.2. When the field is absent, the magnetic moments (the torque exerted on a magnet within a magnetic field) are anti-parallel (Figure 7.2a). Application of the magnetic field turns the magnetic moments through different angles and makes them perpendicular (Figure 7.2b). Finally, it aligns the magnetic moments rendering them parallel (Figure 7.2c). The field saturates the magnetization (the quantity of magnetic moment per unit volume) of the multilayer and leads to a decrease in the electrical
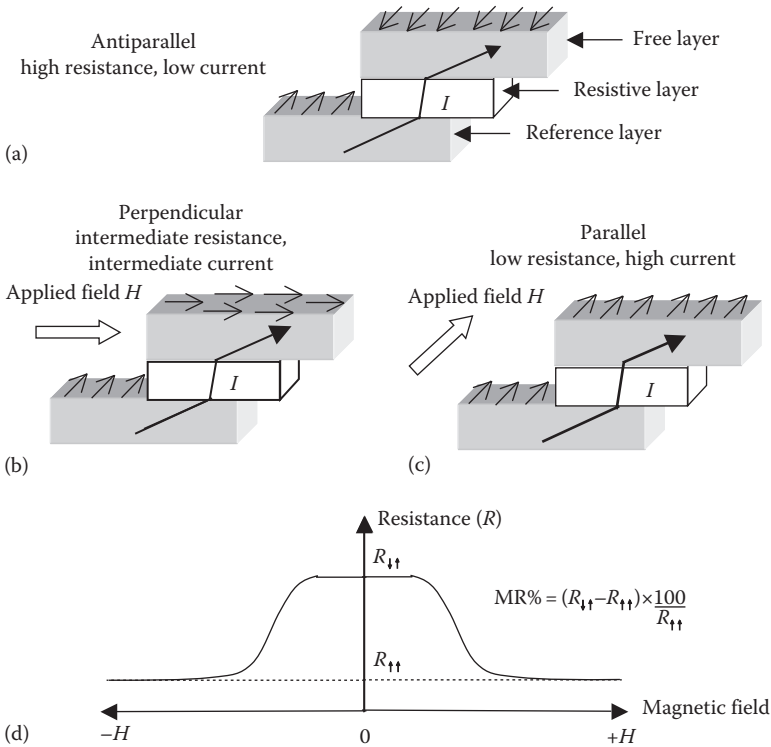
FIGURE 7.2
(a), (b), and (c) Resistance variation with the degree of alignment of magnetic layers sandwiching a non-magnetic layer. The magnetization of the reference layer is pinned and independent of the applied field. The free layer rotates and aligns itself with the applied tickling field H. (d) Typical variation of electrical resistance of the device with magnetic field. $R_{\downarrow\uparrow}$ and $R_{\uparrow\uparrow}$ are resistances measured with antiparallel and parallel magnetizations, respectively. Resistances for antiparallel configuration are much higher than for parallel configuration.

resistance of the multilayer; this resistance variation is depicted in Figure 7.2d. The enormous variation of the resistance is attributed to the scattering of the electrons at the interfaces between the layers.

7.2.3.1 *Scientific Explanation of GMR*

Electrons have the basic properties of mass and charge. The negative charge of the electron is a familiar property, as is its small mass. These properties enable it to move easily through a metal. *In addition to charge and mass, do electrons have a third property that is often overlooked?* Yes, it the *spin* of electrons. It has become possible to take advantage of the spin of electrons for controlling their motion through thin metallic conductors. Electrons spin about their axes, like a child's top. The spin of an electron creates a tiny magnetic field that makes the electron act like a minute magnet.

How many types of spin motion are possible? Spin motion is of two types: up spin (Figure 7.3a) and down spin (Figure 7.3b). In the presence of a magnetic field, the laws of quantum mechanics affirm that an electron can spin in one of two directions: Either its own magnetic field, resulting from its spin motion, aligns with the external magnetic field, which physicists call the "up direction" or its own field aligns opposite to the external magnetic field, which is referred to as the "down direction." In a magnetic material, electron spins are unequal, with the majority of electrons spinning in one direction (Figure 7.3c). Most electrons are aligned in the direction of the cumulative magnetic field, that is, there are more up electrons than down electrons. In a nonmagnetic material, the electron spins are opposite; the up and down electrons counterpoise each other (Figure 7.3d).

In a metal conductor, electricity is transported in the form of electrons, which move freely through the material (Figure 7.4a). The current is conducted because of the movement of electrons in a specific direction. The straighter is the path of the electrons, the greater the conductance of the material. Electrons carrying a current are like balls rolling down a hill. If the slope is smooth, then the balls descend very fast. But if the sloppy hill has a lot of bumps, the balls are slowed down. Defects in the atomic lattice are

(a)　　　　　(b)

(c)　　　　　(d)
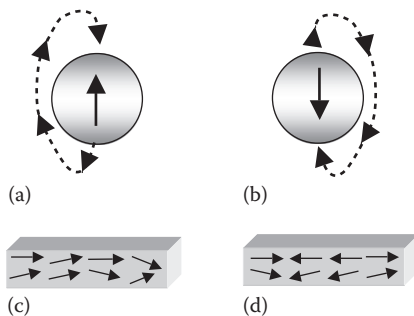
FIGURE 7.3
Spins and materials: (a) up spin, (b) down spin, (c) magnetic material, and (d) nonmagnetic material.

(a)

(b)
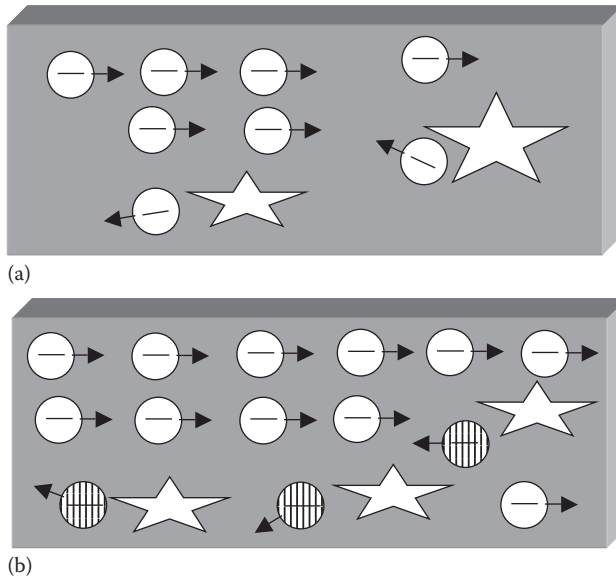
FIGURE 7.4
(a) The electrical resistance in a conductor arises when electrons scatter against irregularities in the material so that their forward movement is obstructed. (b) In a magnetic conductor the direction of spin of most electrons is parallel with the magnetization (empty circles with minus sign). A minority of electrons have spin in the opposite direction (vertically hatched circles with minus sign). Electrons with antiparallel spin are scattered more (http://nobelprize.org/nobel_prizes/physics/laureates/2007/info.pdf).

like these bumps. When electrons collide with these defects, a process that physicists call "scattering," the electrons are slowed down. Therefore, scattering generates electrical resistance. Electric resistance is due to electrons diverging from their straight paths when they scatter against irregularities and impurities in the material. The more the electrons scatter, the higher the resistance is. The resistance of metals depends on the *mean free path* (the average distance covered between successive collisions) of their conduction electrons, which, in GMR devices, depends on the *spin orientation*.

The above must be emphasized. The quantum nature of electrons, that is, either spin up or spin down, is exploited in GMR sensors. In a magnetic material, most of the spins point in the same direction (in parallel), Figure 7.4b. A smaller number of spins, however, always point in the opposite direction, antiparallel to the general magnetization. This spin imbalance gives rise not only to the magnetization as such, but also to the fact that electrons with different spins are scattered to a smaller or greater degree against irregularities and impurities, and especially at the interfaces between materials.

GMR is qualitatively understood using the *Mott model* according to which the electrical conductivity in metals is described in terms of two largely independent conducting channels, corresponding to the up-spin and down-spin

electrons. Unfortunately, there is no way to measure the electrical current carried independently by two kinds of electrons; only the sum of the two currents is measured.

In ferromagnetic materials, conduction electrons either spin up when their spin is parallel to the magnetic moment of the ferromagnet, or spin down when they are antiparallel. The scattering rates of the up-spin and down-spin electrons are quite different, irrespective of the nature of the scattering centers.

In nonmagnetic conductors, there are equal numbers of spin-up and spin-down electrons in all energy bands. Because of the ferromagnetic exchange interaction, there is a difference between the number of spin-up and spin-down electrons in the conduction bands. Quantum mechanics dictates that the probability of an electron being scattered when it passes into a ferromagnetic conductor depends on the direction of its spin. Conduction electrons with spin parallel to the magnetization of the material (spin up) undergo free movement, while the motion of those electrons with antiparallel orientation (spin down) is impeded by collisions with atoms in the material. In general, electrons with spin aligned with the majority of spins in the ferromagnets traverse a longer distance without being scattered. In other words, the scattering is strong for electrons with spin antiparallel to the magnetization direction, and is weak for electrons with spin parallel to the magnetization direction.

1. If the magnetizations are oppositely directed, then electrons originating in one layer are stopped by the adjacent layer (Figure 7.5a). For the antiparallel-aligned multilayer, both the up-spin and down-spin
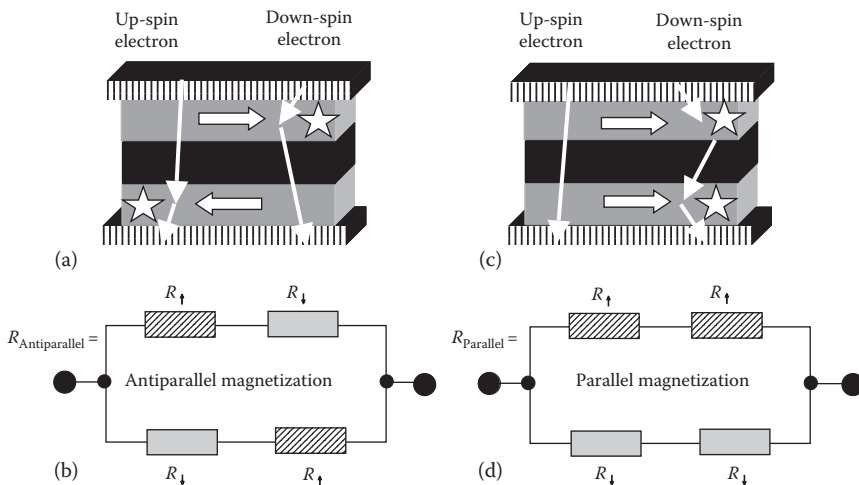
FIGURE 7.5
(a) and (b) Electron transport and equivalent circuit for antiparallel magnetization. (c) and (d) The same for parallel magnetization. CPP case only is considered.

electrons are scattered strongly within one of the ferromagnetic layers, because within one of the layers, their spin is antiparallel to the magnetization direction. The disruption of the free motion of the electrons results in an increase in the electrical resistance. Therefore, in this case, the total resistivity of the multilayer is high. Thus, both electron spins experience small resistance in one layer and large resistance in the other, and the total resistance is

$$R_{\text{Antiparallel}} = \left(\frac{1}{2}\right)(R_\uparrow + R_\downarrow) \tag{7.3}$$

The interfaces in a magnetic multilayer play an important role in spin-dependent transport in bulk elemental ferromagnets. If the interface separates ferromagnetic and nonmagnetic metals, the transmission is spin-dependent due to the spin dependence of the band structure of the ferromagnetic layer, for example, there is a relatively large band mismatch between Cu and the minority spins in Co and thus the transmission of the minority spin electrons across the Co/Cu interface is anticipated to be poor. When the excess spins in the two ferromagnetic layers have opposite directions, both the up- and down-spin electrons are scattered at one of the interfaces. The electrical resistivity for the aligned case is lower because up-spin electrons in this case experience very little resistance and act like a short circuit. Let us visualize the limiting case in which up-spin electrons for the aligned case have no resistance, and down-spin electrons have resistance $2R$, with R being the contribution from each interface. Because the up-spin provides a short-circuit channel, the resistance of the whole system is zero. For the case in which excess spins on two ferromagnetic layers are aligned in opposite directions, both up and down spins have resistance R because both scatter off one interface. In this case, the total resistance is $R/2$, which is definitely greater than the value zero for the aligned case.

2. For the parallel-aligned magnetic layers, that is, when the magnetic layers are aligned in the same direction, the up-spin electrons originating in one layer cross relatively freely through the structure almost without scattering, because their spin is parallel to the magnetization of the layers, as illustrated in Figure 7.5b. On the contrary, the down-spin electrons are scattered strongly within both ferromagnetic layers, because their spin is antiparallel to the magnetization of the two layers. Since conduction occurs in parallel for the two spin channels, the total resistivity of the multilayer is determined mainly by the highly conducting up-spin electrons and appears

to be low. Thus, up-spin electrons experience small resistance, the down-spin electrons experience large resistance, and the sum-total resistance is given by

$$R_{Parallel} = \frac{2R_\uparrow R_\downarrow}{R_\uparrow + R_\downarrow} \tag{7.4}$$

There are similarities between the band structure of Cu and the band structure of the majority spins in Co. This good band matching implies a high transmission probability for the majority-spin electrons across the Co/Cu interface. The up-spin electrons barely notice any difference in the number of electrons per atom in their journey from the ferromagnetic layer to the nonmagnetic layer. To them, the lattice potential is smooth and more or less defect free. On the other hand, the down-spin electrons see a large difference in electron numbers between atoms of copper and cobalt. They perceive several bumps at the interface (because copper and cobalt atoms often intermix there), so they very likely scatter there. When excess spins in the two ferromagnetic layers have the same direction, up-spin electrons travel freely from one ferromagnetic (cobalt) layer across the interface to the other ferromagnetic layer, and down-spin electrons scatter strongly at both interfaces.

From Equations 7.2 and 7.3, the difference between parallel and antiparallel resistances is given by

$$\Delta R = R_{Parallel} - R_{Antiparallel} = -\frac{1}{2} \frac{\left(R_\uparrow - R_\downarrow\right)^2}{\left(R_\uparrow + R_\downarrow\right)} \tag{7.5}$$

Essentially, the interfaces of the Co/Cu multilayer act as *polarizing filters* for the spin of the electrons. When the filters are aligned, the majority-spin electrons pass through relatively easily. When the filters are not aligned, the electrons in both spin channels are reflected at one of the interfaces. To obtain the GMR effect, the spacer layers must be thin compared to the mean free path of electrons to ensure that electrons spin polarized in one layer pass into the other layers before disturbance of their polarization by scattering. The spacer layers allow the magnetic directions of the layers to differ while still permitting the passage of electrons.

From the aforesaid oversimplified arguments, it is evident that resistance is lower when excess spins in the magnetic layers are in the same direction. Thus, if excess spins are pointing in different directions in different magnetic layers when no magnetic field is applied, applying an external magnetic field turns them around so that they point in the same direction, resulting in a decrease in resistivity.

Example 7.2

An electrical resistor implemented with GMR material showed a nominal resistance of 5000 Ω. Then a magnetic field was applied and with this field a minimum resistance of 4400 Ω was achieved. Find (a) MR ratio, (b) the up-spin resistance $R_\uparrow$, and (c) down-spin resistance $R_\downarrow$. (d) Verify that the obtained values satisfy Equation 7.5.

(a) Applying Equation 7.1,

$$MR\% = \left\{ \frac{R_{max} - R_{min}}{R_{min}} \right\} \times 100 = \left\{ \frac{5000 - 4400}{4400} \right\} \times 100 = 13.64\% \qquad (7.6)$$

(b) and (c) Here, $R_{Antiparallel} = 5000$ Ω so that Equation 7.3 gives

$$5000 = \left(\frac{1}{2} \right) \left(R_\uparrow + R_\downarrow \right) \qquad (7.7)$$

$R_{Parallel} = 4400$ Ω, so that Equation 7.4 yields

$$4400 = \frac{2R_\uparrow R_\downarrow}{R_\uparrow + R_\downarrow} \qquad (7.8)$$

Combining Equations 7.7 and 7.8, we have

$$4400 = \frac{2(10,000 - R_\downarrow)R_\downarrow}{10,000 - R_\downarrow + R_\downarrow} = \frac{20,000R_\downarrow - 2R_\downarrow{}^2}{10,000} \qquad (7.9)$$

or

$$2R_\downarrow{}^2 - 20,000R_\downarrow + 44,000,000 = 0 \qquad (7.10)$$

giving

$$R_\downarrow = \frac{20,000 \pm \sqrt{(20,000)^2 - 4 \times 2 \times 44,000,000}}{2 \times 2}$$

$$= \frac{20,000 \pm \sqrt{48,000,000}}{2 \times 2} = \frac{20,000 \pm 6,928.203}{2 \times 2}$$

$$= 6,732.05 \text{ or } 3,267.95 \, \Omega \qquad (7.11)$$

Therefore, the down-spin resistance $R_\downarrow = 6732.05$ or, $3267.95\ \Omega$ and the up-spin resistance $R_\uparrow = 3267.95$ or, $6732.05\ \Omega$.

(d) Putting one set of values of $R_\downarrow = 6732.05\ \Omega$ and $R_\uparrow = 3267.95\ \Omega$ in Equation 7.5,

$$\Delta R = -\frac{1}{2}\frac{\left(R_\uparrow - R_\downarrow\right)^2}{\left(R_\uparrow + R_\downarrow\right)} = -\frac{1}{2}\frac{\left(3{,}267.95 - 6{,}732.05\right)^2}{\left(6{,}732.05 + 3{,}267.95\right)}$$

$$= -\frac{11{,}999{,}988.81}{20{,}000} = -599.99\ \Omega \approx -600\ \Omega \qquad (7.12)$$

The negative value is obtained because ΔR is defined as the difference between R_{Parallel} and $R_{\text{Antiparallel}}$ and $R_{\text{Antiparallel}}$ is $>R_{\text{Parallel}}$. Substituting the other set of values for $R_\downarrow = 3267.95\ \Omega$ and $R_\uparrow = 6732.05\ \Omega$, it is seen that $\Delta R = -600\ \Omega$.

7.2.3.2 Simple Analogies of GMR

To understand how GMR works on the atomic level, consider the following analogies: let us consider a ball to be analogous to a conduction electron. If a person throws a ball between two sets of rollers turning in the same direction (like parallel spin-aligned magnetic layers), the ball tends to go through smoothly. But if the rollers turn in opposite directions, the ball tends to bounce and scatter. Alternatively, the GMR effect may be compared to light passing through optical polarizers. When the polarizers are aligned, light passes through them easily but when their optical axes are rotated with respect to each other, light is prevented from passing across.

7.2.3.3 Optimizing Parameters

What are the critical structural parameters of a GMR device that must have optimum values? Optimal layer thicknesses enhance magnetic-layer antiparallel coupling, which is necessary to keep the sensor in the high-resistance state when no field is applied. When an external field overcomes the antiparallel coupling, the moments in the magnetic layers align and reduce the resistance. If the layers do not have the proper thickness, however, the coupling mechanism can annihilate the GMR effect by causing ferromagnetic coupling between the magnetic layers.

For spin-dependent scattering to be a significant part of the total resistance, the layers must be thinner (to a magnitude of several nanometers) than the mean free path of electrons in most spintronic materials. A typical GMR medical sensor has a conducting layer approximately 3 nm thick. For reference, this is less than 10 atomic layers of copper.

7.2.3.4 GMR Sensor Structures

Besides magnetic multilayers, GMR is exhibited by three main structures: spin valves, pseudo spin valves, and granular solids. Features of all the four structures are presented below:

1. *Magnetic multilayers*: These work by an antiferromagnetic (materials in which the magnetism from magnetic atoms or ions oriented in one direction is canceled out by the set of magnetic atoms or ions that are aligned in the reverse direction) interlayer coupling (Figure 7.6a). Although the measured values of GMR are higher in magnetic multilayers, spin valves are more attractive from the point of view of applications, because only small magnetic fields need to be applied to change the resistance value.

2. *Spin-valve (SV) GMR sensor*: It consists of four layers (Figure 7.6b) (Pelegrí et al. 2003, Qian et al. 2003, Gupta et al. 2010): (i) *Free ferromagnetic layer*: The sensing layer (e.g., NiFe). (ii) *Spacer*: Typically made from copper, this is a nonmagnetic layer that separates the
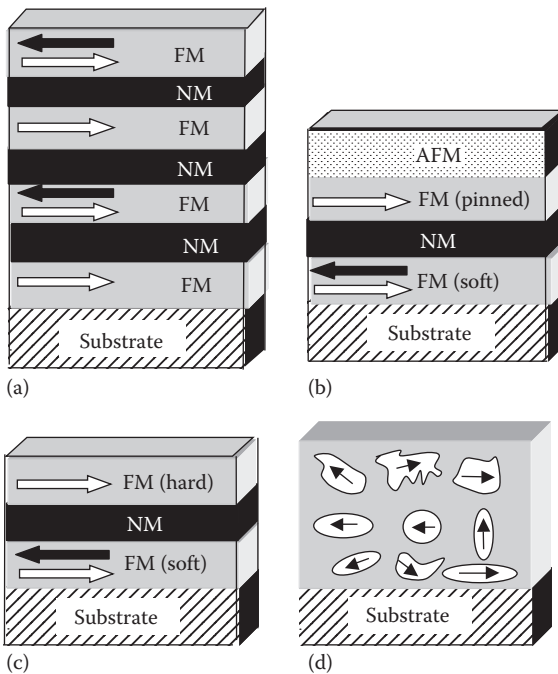
FIGURE 7.6
Different GMR structures: (a) multilayer stack, (b) spin valve, (c) pseudo spin valve, and (d) granular thin film. FM: ferromagnetic, NM: non-magnetic, and AFM: antiferromagnetic.

magnetization of the free and pinned layers. (iii) *Pinned layer*: A layer of cobalt material that is held in a fixed magnetic orientation by its proximity to the exchange layer. (iv) *Exchange layer*: A layer of anti-ferromagnetic material (e.g., FeMn, NiMn, or IrMn) that performs the task of fixing or pinning the magnetic orientation of the pinned layer.

The free layer is sufficiently thin to allow conduction electrons to frequently move back and forth between the free and pinned layers via the conducting spacer layer. The antiferromagnetic layer has no net magnetization of its own, but tends to hold the magneti-zation of the adjacent ferromagnetic layer fixed in direction. Thus, the magnetic orientation of the pinned layer is fixed and held in place by the antiferromagnetic layer, while magnetic orientation of the free layer alters in response to the applied magnetic field. As in other GMR devices, the sensing current flows along the direction of the layers.

This structure has been termed a "spin valve" because one can imagine the magnetic field turning the upper layer like a faucet valve to control the flow of spin-polarized electrons through the device.

3. *Pseudo spin valve*: In this structure, the antiparallel alignment is obtained due to different coercivities of the two ferromagnetic layers (Figure 7.6c). The magnetic moments of the soft and hard magnetic layers switch at different values of the applied magnetic field pro-viding a range of magnetic fields in which they are antiparallel and the resistance is higher.

4. *Magnetic granular solids*: Here, magnetic precipitates are embedded in a nonmagnetic metallic material (Figure 7.6d) (Arana et al. 2004, 2005). In the absence of the field, the magnetic moments of the gran-ules are in state of random orientation. The magnetic field aligns the moments in a certain direction and thus there is a resistance decline.

7.3 Tunneling Magnetoresistance

TMR structures are similar to spin valves. One dissimilarity is that they uti-lize an ultrathin insulating layer to separate two magnetic layers, in place of a conductor. The thickness of this layer is fairly small in order that electron tunneling takes place. Complete prototype sensors are fabricated using an Al_2O_3 (alumina) barrier with a thickness of 1.3 nm. The two electrodes are made of ferromagnetic metal.

TMR devices use the spin valve arrangement of a pinned magnetic layer and a free magnetic layer. Electrons pass from one layer to the other through the insulator by quantum-mechanical tunneling. It is common to electrically connect multiple TMR devices in series to increase the overall resistance and limit the voltage at each tunnel barrier. Voltages above a few hundred millivolts may damage the thin insulator. The electric current tunnels through or flows in a direction perpendicular to the layers (CPP) with contacts on top and bottom of the film stack.

Depending on the orientation of their magnetization, the resistance of the junction varies. The tunneling magnetoresistance (TMR) is the relative variation of the resistance, depending on the magnetization of the two electrodes.

The ease of tunneling between the two magnetic layers is modulated by the angle between the magnetization vectors in the two layers. The two possible configurations are called parallel (P) and antiparallel (AP) configuration. When the magnetization of the layers is aligned, many states are available in the bottom layer into which the spin-polarized electrons from the top layer can tunnel. When the magnetization directions are opposite, the spin-polarized electrons are prevented from tunneling because they do not have the correct orientation for entry in the bottom layer. The process is also known as *spin-dependent tunneling* (*SDT*) (Kasatkin et al. 2000, Freitas et al. 2007).

Figure 7.7 illustrates the configuration that exhibits switchable TMR (Cockburn 2004). A fixed ferromagnet plate is separated from a free ferromagnet plate by a very thin (e.g., ≤1.5 nm thick film) insulating barrier. The magnitude of the tunneling current through the barrier is proportional to the product of the densities of spin-aligned electron quantum states in the conduction sub-band of the plates on either side of the barrier. If the magnetizations (and hence the majority electron spins) in the two plates are parallel, then the majority-spin-aligned electrons in the two plates tunnel more readily across the tunneling barrier. If the magnetizations (and hence
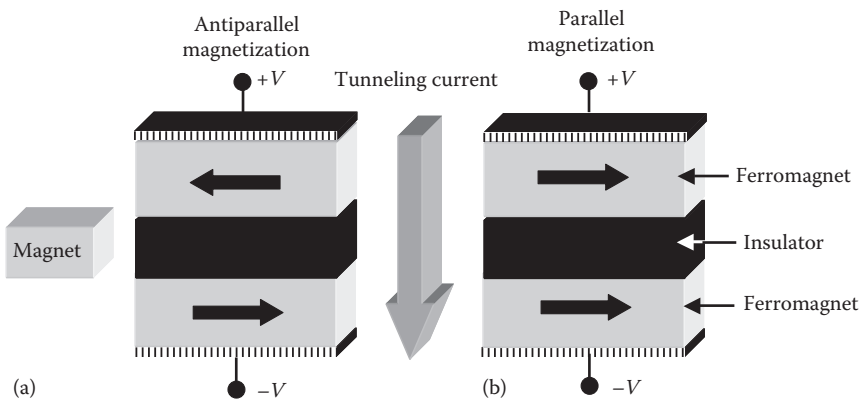
FIGURE 7.7
Tunneling magnetoresistance: (a) antiparallel and (b) parallel spins.

majority electron spins) in the two plates are antiparallel, then the majority-spin-aligned electrons in one plate are inhibited from tunneling to the other plate (and vice versa) because the spin-aligned electron states are in minority there.

Magnetic tunnel junctions are made with transition-metal ferromagnetic materials and exhibit a TMR ratio up to 50% at room temperature. Changes of resistance with magnetic field of up to 70% have been observed in TMR structures. The field required for maximum change in resistance depends upon the composition of the magnetic layers and the method of achieving antiparallel alignment. Values of saturation field range from 0.1 to $10\,\text{kA}\,\text{m}^{-1}$ (1.25–125 Oe) offering at the low end, the possibility of extremely sensitive magnetic sensors. Oersted is defined as $1000/4\pi$ (≈ 79.5774715) A m^{-1}.

7.4 Limitations, Advantages, and Applications of GMR and TMR Sensors

7.4.1 Shortcomings

GMR and TMR sensors are nonlinear and this drawback limits their use as a replacement of Hall sensors for a large number of applications where sensitivity is not crucial. Design of the sensor and the supporting feedback electronics must be done to overcome this problem. *What is the major constraint for using GMR and TMR sensors for accurate measurement applications*? It is the building of hysteresis-free spin valves which is necessary to give an efficient immunity against large random magnetic fields and insure reproducibility of the measurements. Hysteresis is the lagging of an effect behind its cause.

What are other constraints? A second constraint is the linearity of the sensor in the working region. Finally, the reduction of noise, and in particular the magnetic noise, is essential for the use of GMR and TMR as sensors. At low frequencies, the inverse of frequency ($1/f$) noise dominates the thermal noise; it is related to structural defects and magnetic configuration. $1/f$ noise (occasionally called flicker noise or pink noise) is a type of noise whose power spectra $P(f)$ as a function of the frequency f behaves like: $P(f) = 1/f^a$, where the exponent a is very close to 1. Thermal noise is the electronic noise generated by the thermal agitation of the charge carriers, usually the electrons in a conductor.

7.4.2 Advantages

Notwithstanding their drawbacks, compared to Hall effect and AMR sensors, the manifold advantages of GMR sensors must be mentioned: (i) They allow for large magnetic field–dependent changes in resistance, a larger

output signal and the ability to have a larger air gap (distance from the sensor to the target), enabling tighter mechanical tolerance in system components to be achieved. (ii) GMR devices are, apparently, not susceptible to damage from large magnetic fields and also do not tend to exhibit any latch-up effect. (iii) Dense packing of magnetic sensors in small areas is carried out using photolithography for TMR devices with high resistance measuring only several micrometers on a side. Extremely small SDT devices, several μm on a side, with high resistance values are fabricated using photolithography allowing very dense packing of magnetic sensors in small areas. (iv) SDT magnetic sensors have a number of attractive properties such as a pronounced spin-tunneling effect, sensitivity at the picotesla level (10–12 tesla; tesla is the SI unit of magnetic field: a particle carrying a charge of 1 C and passing through a magnetic field of 1 T at a speed of 1 m s⁻¹ experiences a force of 1 N), high resistance and small size of the junction, and low density of the sensor current (Kasatkin et al. 2000). (v) From comparison with AMR sensors using similar materials, one should expect wide operational temperature and frequency ranges for SDT sensors.

7.4.3 Applications

The largest technological application of GMR is in the *data storage industry*. Disk drives based on GMR head technology use their properties to control a sensor that responds to very small magnetic rotations on the disk. The magnetic rotation yields a very large change in sensor resistance, which in turn provides a signal that is collected by the electric circuits in the drive. Applications of GMR are diverse, for example, in automotive sensors, solid-state compasses and nonvolatile magnetic memories (nonvolatile means that the memory is maintained even when power to the device is switched off).

Focusing on the hard disk, the GMR read-head sensor in a hard disk is manufactured using a spin valve structure (Figure 7.8). Spin valve resistance shows a steep change in the small field range around $H = 0$. As the magnetic bits on the hard drive pass under the read head, the magnetic alignment of the sensing layer in the spin valve changes resulting in variation of its resistance. When the head passes over a magnetic field of one polarity, the electrons in the free layer respond by aligning with those on the pinned layer, creating a lower resistance in the head structure. When the head passes over a field of opposite polarity, the free layer electrons rotate so that they are no longer aligned with the electrons in the pinned layer. This increases the resistance of the structure. Because the resistance changes are ushered by changes to the spin characteristics of electrons in the free layer, GMR heads are also known as *spin valves*, a term that was coined by IBM.

What are other important applications of GMR sensors? Besides data storage, GMR sensors are also widely used in biological analysis for biosensors that measure the presence of DNA or antibodies, immunosorbent assays (Millen et al. 2005), magnetic arrays, and other applications. An immunosorbent
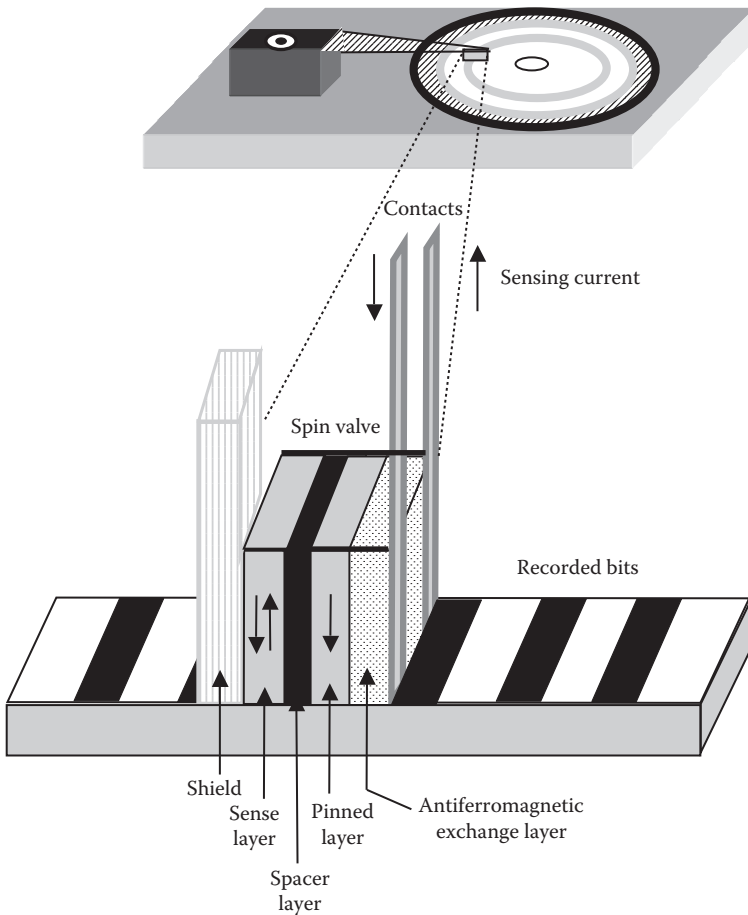
FIGURE 7.8
Application of GMR spin valve sensor in a hard disk as a reading device.

assay is a biochemical technique used in immunology to detect the presence of an antibody or an antigen in a sample.

7.5 Magnetic Nanoparticle Probes for Studying Molecular Interactions

Research on nanoparticles of metals and semiconductors has been extensively pursued worldwide because of their unique electronic and optical properties. When coupled to affinity ligands, such nanoparticles function

as sensitive biological nanosensors. A ligand is a substance that forms a complex with a biomolecule to serve a biological purpose. The interaction of ligands with their binding sites is characterized in terms of a binding affinity. In general, high-affinity ligand binding results from greater intermolecular force between the ligand and its receptor while low-affinity ligand binding involves smaller intermolecular force between the ligand and its receptor.

Alternative detection technologies have been developed using magnetic nanosensors composed of magnetic nanoparticles that are used to detect molecular interactions by magnetic resonance (MR) such as nuclear magnetic resonance (NMR)/magnetic resonance imaging (MRI) techniques.

What do NMR/MRI mean? NMR is a medical imaging method to visualize body structures and their functioning, in which the nuclei of material placed in a strong magnetic field absorb radio waves supplied by a transmitter at particular frequencies. MRI is a diagnostic imaging technique that applies a magnetic field and radiowaves to produce proton density images from NMR of protons.

What is the composition, structure, and dimensions of magnetic nanosensors? Magnetic nanosensors are composed of 3–5 nm monocrystalline iron oxide nanoparticles (MION), with an inverse spinel structure (cubic close packed) of $(Fe_2O_3)_n(Fe_3O_4)_m$. Owing to their magnetic properties, iron oxide nanocores have a natural tendency to aggregate in suspensions.

What methods are adopted to prevent the particles from self-aggregation? To keep particles separated, two principal ways of particle stabilization have been developed. (i) The first option is *steric stabilization*, where particles are coated with a thick layer of macromolecules that keep the iron oxide cores sterically apart. Dextran (complex, branched glucan [polysaccharide made of many glucose molecules]) and polyethylene glycol (PEG) polymers are extensively used for nonspecific and molecular imaging probes. These polymer coatings allow covalent coupling of chemical compounds, peptides (Schellenberger et al. 2004, Schellenberger 2010), and proteins without impairing their sterical stabilization. Therefore, sterically stabilized superparamagnetic particles have dominated molecular probe development for MRI. "Steric" relates to the spatial arrangement of atoms in a molecule. Steric effects originate from the fact that each atom within a molecule occupies a fixed amount of space. If atoms are brought too close together, there is an associated expenditure of energy due to overlapping electron clouds, and this affects the preferred shape (conformation) and reactivity of the molecule. (ii) The second option is electrostatic stabilization with positively or negatively charged surface coatings. The advantage of electrostatic stabilization over steric method is the possibility of developing smaller specific superparamagnetic nanoparticles with similar contrast effect. However, direct modification of the stabilizing charged surface frequently results in an alteration of the surface charge inducing destabilization and particle aggregation.

What is the average size of these nanoparticles? Depending on their preparation, the average size of the nanoparticles surrounded by a 10 kDa dextran

(a glucose polymer) coating of approximately 10 nm thickness is about 25–30 nm (Perez et al. 2004). This is equivalent to a 750–1200 kDa globular protein (protein that consists of long chains of amino acids folded up into complex shapes). 1 kDa = the weight of 1000 hydrogen atoms = 1.66×10^{-21} g.

Are these nanoparticles superparamagnetic? Yes.

What is superparamagnetism? Superparamagnetism occurs when a material is composed of very small crystallites (1–10 nm) (Gossuin et al. 2009). In this case, even though the temperature is below the Curie or Neel temperature (the temperature above which a ferromagnetic substance loses its ferromagnetism and becomes paramagnetic) and the thermal energy is not sufficient to overcome the coupling forces between neighboring atoms, the thermal energy is adequate to change the direction of magnetization of the entire crystallite. The resulting fluctuations in the direction of magnetization cause the magnetic field to average to zero. The material behaves in a manner similar to a paramagnetic material, except that instead of each individual atom being independently influenced by an external magnetic field, the magnetic moment of the entire crystallite tends to align with the magnetic field. Essentially, superparamagnetism is a form of magnetism that occurs only in the presence of an externally applied magnetic field.

The nanoparticles become magnetized when placed in an external magnetic field and lose their magnetic moment when the field is withdrawn (Haun et al. 2010). This phenomenon arises because the nanoparticles consist of a single crystal domain and thus exhibit a net magnetic moment directed along an anisotropy axis. Aided by thermal energy, however, the magnetic moment can overcome the anisotropy barrier and spontaneously flip from one direction of anisotropy to another. An ensemble of magnetic nanoparticles consequently displays negligible magnetic remanence. The remanence is the strength of the magnetic field that remains in a magnetic nanoparticle after it is exposed to a strong, external magnetic field and the external field is removed.

When placed inside an external magnetic field, the magnetic moments of the above-mentioned nanoparticles align in the direction of the field lines and enhance the magnetic flux. This behavior is similar to paramagnetism, but since it is associated with a large particle with fixed magnetic moment, it is referred to as superparamagnetism.

How are neighboring water protons affected when these particles are placed in a magnetic field? When subjected to an external magnetic field, each superparamagnetic nanoparticle produces a large magnetic dipole. The resulting local magnetic field gradient creates an inhomogeneity in the external magnetic field, which destroys the coherent precession of nuclear spins of neighboring water protons.

What is the life of magnetic nanoparticles? Different MION preparations have a blood half-life of >10 h in mice. Related clinical nanoparticle preparations have circulation times of 24 h in humans.

What are amino-CLIO nanoparticles? CLIO (cross-linked iron oxide) nanoparticles (Haun et al. 2010) contain a superparamagnetic iron oxide core (3–5 nm sized MION) composed of ferromagnetic magnetite (Fe_3O_4) and/or maghemite (γ-Fe_2O_3). Maghemite is structurally and functionally similar to magnetite but differs in one aspect in that it contains cation vacancies in the sublattice.

How are functional groups attached to the magnetic nanoparticles? To develop more stably coated and amino-functionalized sensors, the iron oxide core is coated with dextran. The dextran coating is cross-linked with epichlorohydrin (C_3H_5ClO: an organochlorine compound and an epoxide). It is successively treated with epichlorohydrin to form stabilizing cross-links and ammonia (NH_3) to provide primary amine group functionality. Preparations of the resulting aminated cross-linked iron oxide nanoparticle (amino-CLIO) have about 40 amino groups per particle with an average particle size of 40–50 nm.

Amino groups in amino-CLIO react by *N*-hydroxysuccinimide (NHS), $C_4H_5NO_3$-based bifunctional (having two functions) cross-linking (a covalent bond is formed between polymer chains, either within or across chains). This allows the attachment of a range of sulfhydryl (-C-SH or R-SH where R represents an alkane, alkene, or other carbon-containing segment)-bearing biomolecules, giving rise to biomolecule–nanoparticle conjugates with unique biological properties as compared to simple polymer-coated iron oxide nanoparticles.

What kind of treatments or handling can these coated magnetic nanoparticles forebear? These nanoparticles can withstand harsh treatments, such as incubation (maintaining at a particular temperature for a set length of time) at 120°C for 30 min, without any change in their size or loss of their dextran coating.

How are molecular interaction probing experiments performed with magnetic nanoparticles? Binding of these magnetic nanoparticles with their intended molecular targets, leads to the formation of stable nanoassemblies (Figure 7.8). As a result, there is a corresponding decrease in the spin–spin relaxation time (T_2) of surrounding water molecules. This relaxation time T_2 of water molecules is the primary parameter of attention in these studies. Thus, the core thought behind the use of magnetic nanoparticles is to employ them as *proximity sensors* that modulate the spin–spin relaxation time of neighboring water molecules. This relaxation time is quantified using common clinical MRI scanners or benchtop NMR relaxometers.

What are the salient features of MRI scanners and benchtop relaxometers? MRI scanners have the disadvantages of (i) high operating costs, (ii) bulky equipment size (mainly due to superconducting magnets: electromagnets made from coils of superconducting wires), and (iii) the need for relatively large sample volumes (hundreds of microliters). On the other hand, benchtop relaxometers provide a lower cost alternative to MRI scanners, making them more accessible for use in diagnostic MR sensing. Benchtop systems operate

at lower NMR frequencies (100 kHz–50 MHz) and are equipped with a permanent, low-field (<1 T) magnet for field generation. However, benchtop systems have two drawbacks: (i) they lack the capability to perform parallel measurements and (ii) they still require relatively large sample volumes.

It is necessary to recall here: *What is magnetic relaxation?* It is reiterated that magnetism is associated with angular momentum called *spin*, because it usually arises from spin of nuclei or electrons. The spins may interact with applied magnetic fields, the so-called *Zeeman energy* (the energy of the interaction between the magnetic moment of an atom or molecule and an applied magnetic field); with electric fields, usually atomic in origin; and with one another through *magnetic dipole or exchange coupling* (the coupling between two nuclear spins due to the influence of bonding electrons on the magnetic field running between the two nuclei), the purported *spin–spin energy*. *Magnetic relaxation* phenomenon is defined as the approach of a magnetic system to an equilibrium or steady-state condition, over a period of time. Relaxation involves several processes by which nuclear magnetization created in a nonequilibrium state reverts to the equilibrium distribution. Further, the relaxation is not an instantaneous phenomenon but requires time. The characteristic times involved in magnetic relaxation are known as *relaxation times*.

What are the different relaxation processes and corresponding relaxation times? Different physical processes are responsible for the relaxation of the components of the nuclear spin magnetization vector **M** parallel and perpendicular to the external magnetic field, $\mathbf{B}_0$ (which is conventionally oriented along the z axis). These two principal relaxation processes are designated as T_1 and T_2 relaxation, respectively.

What is spin-lattice relaxation? It is the magnetic relaxation in which the excess potential energy associated with electron spins in a magnetic field is transferred to the lattice. (As used here, the term lattice does not refer to an ordered crystal but instead signifies degrees of freedom other than spin orientation, for example, translational motion of molecules in a liquid.). *What is spin-lattice relaxation time T_1?* T_1, also called *longitudinal relaxation time*, characterizes the rate at which the longitudinal M_z component of the magnetization vector recovers. It is thus the time taken by the signal to recuperate around 63% [1 – (1/e)] of its starting value after being flipped into the magnetic transverse plane.

What is spin–spin relaxation? It is the magnetic relaxation, observed after application of a weak magnetic field, in which the excess potential energy associated with electron spins in a magnetic field is redistributed among the spins, resulting in heating of the spin system. It is a complex phenomenon corresponding to a decoherence or dephasing of the transverse nuclear spin magnetization. *What is spin–spin relaxation time?* Spin–spin relaxation time, also known as *transverse relaxation time* (T_2), is a time constant in NMR and MRI. It is named in contrast to T_1, the spin-lattice relaxation time. T_2 characterizes the rate at which the M_{xy} component of the magnetization vector decays in

the transverse magnetic plane. It is the time taken by the transverse signal to fall to 37% ($1/e$) of its initial value after flipping into the magnetic transverse plane. Hence, the following formula describes the process:

$$M_{xy}(t) = M_{xy}(0)\exp\left(\frac{-t}{T_2}\right) \tag{7.13}$$

T_2 decay occurs 5–10 times more rapidly than T_1 recovery, and different tissues have different values of T_2. Fluids have the longest $T_2 \sim 700$–$1200\,\text{ms}$. Water-based tissues have T_2 in the 40–200 ms range while fat-based tissues have T_2 in the 10–100 ms range.

It may be noted here that spin–spin relaxation does not change the total energy of the spin interactions whereas spin-lattice relaxation modifies the same. Further, spin–spin relaxation is associated with an internal equilibrium of the spins amongst themselves. Spin-lattice relaxation is associated with the approach of the spin system to thermal equilibrium with the host material.

Returning and referring to the discussions on molecular interaction probing experiments, *what kind of assays are carried out with magnetic nanoparticles?* Sensitive homogenous assays are performed using magnetic nanoparticles to detect a variety of different molecular interactions in biological samples with minimal or no sample preparation. These magnetic nanosensors are able to detect specific mRNA (messenger RNA), proteins, enzymatic activity, and pathogens (e.g., virus) with sensitivity in the low femto molar range (0.5–30 fmol).

What is the mechanism of magnetic relaxation switching in the above assay experiment? The iron oxide crystal core in CLIO being superparamagnetic becomes magnetized when placed in an external magnetic field. The combined electron spins in the crystal produce a single large magnetic dipole, creating a local magnetic-field gradient and therefore an inhomogeneity in the external magnetic field. Water protons diffusing within the local inhomogeneity process at an off-resonance frequency, dephasing their spins, thus increasing the relaxation rate ($1/T_2$) as described by outer-sphere theory, in which $1/T_2$ is directly proportional to nanoparticle cross-sectional area.

Dephasing or decoherence is the process by which quantum-mechanical interference is destroyed so that coherence in a substance caused by perturbation decays over time, and the system returns to the state before perturbation; the mechanism that recovers classical behavior from a quantum system.

Outer-sphere electron transfer is, by definition, intermolecular. Electron transfer is the process by which an electron moves from one atom or molecule to another atom or molecule. Outer sphere electron transfer occurs between identical or dissimilar chemical species differing only in their oxidation state. Here, the participating redox centers are not linked via any
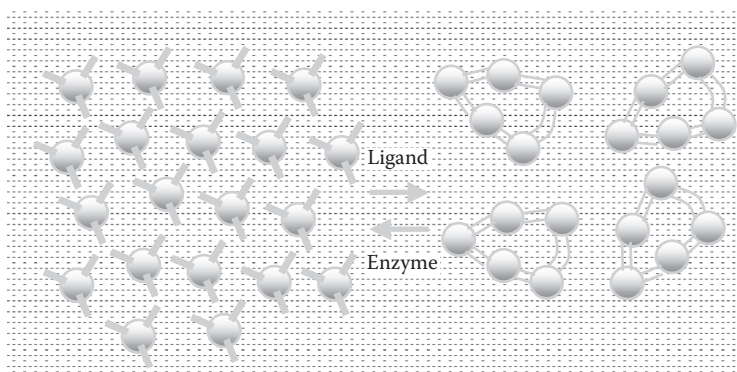
FIGURE 7.9
Self-assembly of superparamagnetic particles in the presence of a target with an associated decrease in T_2 relaxation time.

bridge during the electron event so that the electron hops through space from the reducing center to the acceptor.

It is hypothesized that, when individual superparamagnetic nanoparticles assemble into clusters (Figure 7.9) and the effective cross-sectional area becomes larger, the nanoassembly becomes more efficient at dephasing the spins of surrounding water protons, leading to an enhancement of the relaxation rates ($1/T_2$). Hence, these magnetic nanosenors are referred to as *magnetic relaxation switches*. They act as switches through target-mediated clustering. Due to the presence of molecular targets, they switch from a dispersed state to a clustered state or vice versa, with changes in spin–spin relaxation time of water (T_2).

Interestingly, the spin-lattice relaxation time (T_1) is independent of nanoparticle assembly formation. Hence, this parameter is used to measure concentration in both nano-assembled and dispersed states within the same solution. Nanoassembly formation is designed to be reversible (e.g., by temperature, chemical cleavage, pH, etc.) so that forward or reverse assays can be designed and implemented. Several examples of forward (clustering) or reverse (declustering) types of assays for detecting a large variety of biologically relevant materials can be cited.

Biosensing strategies based on magnetic nanoparticles have received considerable attention because they offer unique advantages over other techniques. *What are the special advantages of this method?* These include the following: (i) Rapid detection of a target without elaborate purification of the sample or requirement of signal amplification. (ii) Biological samples exhibit virtually no magnetic background, and, thus, highly sensitive measurements are performed in turbid or otherwise visually obscured samples without further processing. The use of light as in analysis entailing fluorescence, absorbance, chemiluminescence, etc., is avoided. The significant benefit is that light does not affect the outcome of the assay, and experiments

are conducted in turbid, light-impermeable media such as blood, cell sus-
pensions, culture media, lipid emulsions, and even whole tissue. (iii)
Immobilization of the sample on a flat surface, for example, microarray glass
slides, is not necessary. Hence, faster hybridization and binding kinetics are
observed. (iv) The assay provides the flexibility to detect various biomolecu-
lar interactions like mRNA, protein, and enzymatic activity simultaneously.
It is accomplished in a high-throughput format by using MRI and NMR.
(v) At the low iron concentrations (<20 μg Fe per mL) used in the assay, the
nanometer-sized clusters do not aggregate or precipitate. (vi) Finally, mag-
netic nanoparticles are inexpensive to produce, physically and chemically
stable, biocompatible, and environmentally safe. Since identical iron oxide
nanoparticles are commonly used in clinical studies and have shown little-
to-no toxicity, the assay is applicable to *in vivo sensing* of molecular targets
by MRI.

Example 7.3

For tissues having $T_2 = 85$ ms, find the time in which $M_{xy}(t)$ component of
magnetization decreases by 50%.

Here, $M_{xy}(0) = 100$ and $M_{xy}(t) = 50$. By Equation 7.13,

$$50 = 100 \times \exp\left(\frac{-t}{T_2}\right) \tag{7.14}$$

or

$$\ln(0.5) = \frac{-t}{85} \tag{7.15}$$

from which $-0.69315 = -t/85$; therefore, $t = 58.92$ ms.

7.5.1 DNA Analysis

Josephson et al. (2001) and Perez et al. (2002a,b) conducted several experi-
ments to detect oligonucleotide sequences by coupling an average of three
oligonucleotides (12 base pairs) to the nanoparticles using *N*-succinimidyl-
3-(2-pyridyldithio)propionate (SPDP) as a linker. SPDP (*N*-succinimidyl
3-(2-pyridyldithio)-propionate), $C_{12}H_{12}N_2O_4S_2$, is a heterobifunctional, thiol
(organosulfur compound)-cleavable and membrane permeable crosslinker.
It contains an amine-reactive NHS ester, $C_4H_5NO_3$ that reacts with lysine
($C_6H_{14}N_2O_2$) residues to form a stable covalent amide bond; amine is an
organic compound/functional group that contains a basic nitrogen atom
with a lone pair. The other end of the spacer arm is terminated in the pyridyl
disulfide ($C_{10}H_8N_2S_2$) group that will react with sulfhydryls (–C–SH or R–SH
where R = an alkane, alkene, or C-containing part of molecule) to form a

reversible disulfide bond (covalent bond, usually derived by the coupling of two thiol groups: R–S–S–R). An oligonucleotide is a short nucleic acid polymer of 2–20 nucleotides. Their methodology is as follows: Two separate CLIO nanoparticle biosensors were prepared by attaching different complementary 12 base pair oligonucleotides that recognized adjacent sites on a 24 base pair target sequence, that is, for each intended target sequence, they made two unique nanoparticle populations (termed P_1 and P_2), recognizing adjacent 24 base pair long target sequences. Addition of the complementary target oligonucleotide sequence resulted in a rapid (<30 min) cluster formation and aggregation of these nanoparticles because oligonucleotides readily bind to their respective complementary nucleotides nanoparticle.

Upon hybridization with a target sequence, the particles oligomerized into larger assemblies of approximately 200 nm. This resulted in a quick and significant decrease in the spin–spin relaxation times (T_2) of neighboring water molecules, as measured with a 0.47 T NMR relaxometer at 40°C. When a noncomplementary oligonucleotide was used, no change in T_2 was noticed. The average decrease in T_2 was linear with the amount of DNA added in the concentration range displayed. Similar changes in T_2 relaxation times were observed when the experiments were conducted in turbid solution.

The nanoparticle assemblies could be dissociated by heating. Moreover, no change in T_2 was observed when scrambled oligonucleotide sequences were used, but inclusion of even a single mismatch in the target sequence resulted in a detectable signal change relative to the perfectly matched sequence. This technique may therefore prove valuable for mutational analysis, biochemical identification of mutational changes in a nucleotide sequence. Using this technique, they were successful in detecting as low as 10 fmol of DNA while lower amounts (0.5 fmol) were detected using a 1.5 T MR imager.

7.5.2 Protein Detection

Biosensors were formerly developed to detect soluble proteins by leveraging the specificity of antibodies (proteins produced by the body's immune system when it detects harmful substances, called antigens). The first illustration involved detection of *green fluorescent protein* (GFP) to establish a proof of the principle. GFP from the jellyfish *Aequorea victoria* is a 27 kDa monomer protein consisting of 238 amino acids. It fluoresces green in the presence of UV light. It is a molecular tag that can be inserted into genes to make animals and plants glow green.

GFP-sensitive CLIO was prepared by conjugating with avidin, followed by attachment of biotinylated anti-GFP polyclonal antibody; a polyclonal antibody or antisera is the antibody derived from multiple B cells or cell lines. Perez et al. (2002a,b) prepared avidin-P1 conjugates as generic reagents for attachment of any biotinylated antibody (or peptide) to the nanosensor. Normally, each nanoparticle contained two avidins (i.e., eight binding sites). Biotinylated anti-GFP polyclonal antibody was attached to yield a

GFP-sensitive nanosensor. When the nanosensors were used to investigate GFP protein, noteworthy changes in T_2 relaxation time were found. GFP was speedily (<30 min) detected in a dose-dependent manner to as low as femtomolar range. These changes were time- and dose-dependent. Incubation with control protein (bovine serum albumin [BSA]) showed no major changes in T_2.

In a different study (Kim et al. 2007), a slightly dissimilar scheme was implemented to detect the beta subunit of human chorionic gonadotrophin (HCG-β), a glycoprotein composed of 244 amino acids with a molecular mass of 36.7 kDa, and a biomarker concerned with prostate and ovarian cancers. In this case, two different monoclonal antibodies that bind disparate, nonoverlapping epitopes (localized regions on the surface of an antigen that is capable of eliciting an immune response and of combining with a specific antibody to counter that response) on the HCG-β protein were attached to separate CLIO nanoparticle populations. Monoclonal antibodies are the highly specific antibodies produced in large quantity by the clones of a single hybrid cell formed in the laboratory.

Using this system, HCG-β was convincingly detected in a dose-dependent manner, and the sensitivity improved when an HCG dimmer was used as the target along with increasing the number of antibodies per CLIO nanoparticle.

7.5.3 Virus Detection

Perez et al. (2003) reported the construction of magnetic nanosensors having the capability of detecting complex targets, such as intact viral particles in serum (a component of blood which is collected after coagulation). Since proteins are promptly detected by using this technique, they argued that magnetic nanoparticles coated with antibodies against virus surface proteins could be applied to identify the viral particle in solution. They conjectured that the multiple interactions occurring between a multivalent target (virus) and a multivalent magnetic nanoparticle–antibody conjugate could be utilized in an extremely sensitive assay for the detection of viruses.

The developed magnetic viral nanosensors were composed of a superparamagnetic iron oxide core caged with a dextran coating on which virus-surface-specific antibodies were attached (Figure 7.10). Binding of anti-denovirus 5 (ADV-5) or anti-herpes simplex virus 1 (HSV-1) antibodies was accomplished via protein G coupling, attached to the caged dextran via *N*-succinimidyl- 3-(2-pyridyldithio)propionate (SPDP); the SPDP protein $(C_{12}H_{12}O_4N_2S_2)$ is a cleavable, water insoluble, amino and thiol (sulfhydryl) reactive heterobifunctional protein crosslinker. By and large, the particle size of the magnetic nanoparticles was 46 ± 0.6 nm.

They first explored whether incubation of the ADV-5 viral particles with the anti-ADV-5 magnetic nanosensor will be able to form nanoassemblies in solution. Immediately after addition of 10^4 viral particles to the magnetic nanosensors, two distinct populations of particles were detected by light
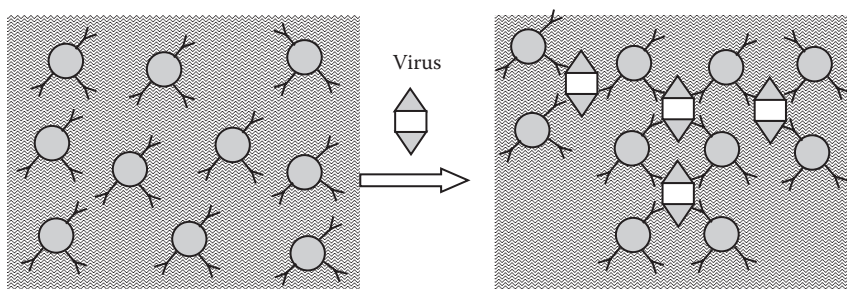
FIGURE 7.10
Magnetic viral nanosensors: clustering of nanoparticles in presence of virus. (After Perez, J. M., et al., *J. Am. Chem. Soc.*, 125, 10192, 2003.)

scattering. These pertained to magnetic nanoparticles alone (46 ± 0.6 nm) and viral particles (100 ± 18 nm). Within 30 min of incubation, the viral particle population became undetectable by light scattering and a larger-size population (494 ± 23 nm) appeared, corresponding to the viral-induced nanoassembly. The size of the nanoassembly continued to increase to some extent, reaching a plateau within 2 h (550 ± 30 nm). Atomic force microscopy (AFM) studies confirmed the presence of viral-induced nanoassemblies in solution. They investigated the effects of nanoassembly formation on the T_2 relaxation times of water at 0.47 T (20 MHz) with a tabletop relaxometer. Within 30 min, ΔT_2 was half-maximum and reached a plateau in less than 2 h.

Using this approach, they were able to detect low levels (five viral particles in 10 µL) of herpes simplex virus-1 (HSV-1) and adenovirus-5 (ADV-5) in serum solutions. The method is also independent of the optical properties of the solution, permitting the detection of the virus in complex turbid media.

7.5.4 Telomerase Activity Analysis

Telomerase is a eukaryotic ribonucleoprotein (RNP) complex containing both an essential RNA and a protein reverse transcriptase subunit. Using its intrinsic RNA as a template for reverse transcription, it helps to stabilize telomere length in human stem cells, reproductive cells, and cancer cells by adding TTAGGG repeats onto the telomeres to the 3′ end of DNA strands in the telomere regions, found at the ends of chromosomes. Telomerase activity has been found in almost all human tumors but not in adjacent normal cells. High levels of telomerase activity are found in most malignancies and are believed to play a critical role in tumorigenesis offering an attractive target for therapeutic intervention and diagnostic or prognostic purposes. This makes telomerase a target not only for cancer diagnosis but also for the development of novel anticancer therapeutic agents. The ability to detect telomerase activity rapidly, quantitatively, and repeatedly would have significant value in cancer research, the development of more efficient

telomerase inhibitors, titrating treatment efficacy, and in the detection of enzyme levels for prognostic (predictive) purposes. Grimm et al. (2004) described the use of magnetic nanoparticles for investigating telomerase activity for rapid screening of telomerase activity in biological samples. The basis is hybridization of magnetic nanoparticles (consisting of caged iron oxide crystals) to telomerase-associated TTAGGG repeats. Due to nanoparticle assembly formation, the relaxation time (T_2) of surrounding water changes significantly, which is readily measured by MR relaxometers or imaging systems.

Amidated cross-linked iron oxide nanoparticles (CLIO-NH$_2$) were conjugated to oligonucleotides via a stable thioether (a functional group in organosulfur chemistry with the connectivity C–S–C) linkage with *N*-succinimidyl iodoacetate (C$_6$H$_6$INO$_4$) a sulfhydryl and amino reactive, heterobifunctional protein cross-linking reagent. Each CLIO particle had an average of four oligonucleotides bound. Aminated CLIO (amino-CLIO) nanoparticles have an average hydrodynamic diameter of 25–40 nm and approximately 40–80 amines available per particle for conjugation of biomolecules. Hydrodynamic diameter of a particle is the effective diameter in a solution measured by assuming that it is a body moving through the solution and resisted by viscosity of the solution. If the solvent is water, the hydrodynamic diameter includes all the water molecules attracted to the particle. As a result, it is possible for a small particle to have a larger hydrodynamic diameter than a large particle, if it is surrounded by more solvent molecules. Simply stated, the hydrodynamic diameter is not measuring the particle, but is effectively measuring the nanoparticle and anything affiliated with its surface, for example, a surrounding double-layer, polymers or other capping agents, or nanoparticle aggregates. Thus, hydrodynamic diameter is an equivalent sphere diameter derived from a measurement technique involving hydrodynamic interaction between the particle and fluid.

The amine groups can then be reacted with various reagents to attach biomolecules via anhydride (formed from another compound by removal of water), amine (a functional group (–NH$_2$) formed by the loss of a hydrogen atom from ammonia), hydroxyl (–OH), carboxyl (–COOH), thiol (–SH), or epoxide (an oxygen atom joined by single bonds to two adjacent carbon atoms, thus forming the three-membered epoxide ring) groups. The resultant sensors were stable in solution over months, were monodisperse (in which the suspended particles have identical size, shape, and interaction) and had an overall size of 45 ± 4 nm. When sensors were incubated with telomeric repeats, they assembled in a linear fashion along the repeats, a feature that was not observed with control nonsense sequences.

The above researchers tested the efficacy of different telomerase inhibitors in crude human and murine (relating to a rodent of the family Muridae or subfamily Murinae, including rats and mice) samples and showed that phosphorylation (the addition of a phosphate (PO$_4$) group to a

protein or other organic molecule) of telomerase regulates its activity. High-throughput adaptation of the technique by MRI allowed processing of hundreds of samples hundreds of samples within tens of minutes at ultrahigh sensitivities.

They reported that all tumor cell lines (breast, prostate, ovarian, pancreatic, lung, and liver carcinoma and glioma, type of tumor that starts in the brain or spine; melanoma, the most dangerous type of skin cancer; lymphoma, a type of cancer involving cells of the immune system, called lymphocytes; and insulinoma, tumor of the pancreas that produces excessive amounts of insulin) and primary tumor tissue samples (liver metastasis) as well as human skin fibroblasts (the most common type of cells found in connective tissues) from newborns tested positive for telomerase, whereas primary skin melanocytes (a pigment-producing cell in the skin that determines its color) tested negative.

7.6 Protease-Specific Nanosensors for MRI

Imaging of enzyme activity is a central goal of molecular imaging. With the introduction of fluorescent smart *in vivo* detection of enzyme activity probes, optical imaging has become the modality of choice for experimental *in vivo* detection of enzyme activity. Perez et al. (2002a,b) introduced magnetic relaxation switches sensing enzymatic activity based on superparamagnetic high-relaxivity probes for *in vitro* MRI. Upon protease activation, these sterically stabilized cross-linked iron oxide particles (CLIO) switch from large clustered high-relaxivity aggregates to smaller low-relaxivity single particles. This technique is not suitable for *in vivo* application, since the nonactivated clustered particles are too large to have a favorable bioavailability and, more importantly, activation by elevated enzyme activity in the target tissue *reduces* the relaxivity and consequently the contrast effect. Schellenberger et al. (2008) described a new design of protease-sensitive nanosensors (Figure 7.11), which are based on electrostatically stabilized, citrate-coated very small iron oxide particles (VSOP) with a hydrodynamic diameter of 7.7 ± 2.1 nm instead of larger sterically stabilized nanoparticles (over 20 nm). Protease is a digestive enzyme that catalyzes the hydrolytic breakdown of proteins into peptides or amino acids.

The synthesized fluorescein ($C_{20}H_{12}O_5$)-labeled MMP-9 (matrix metalloproteinase 9) peptide NH_2-GG*PRQITA*G-K(FITC)-GGGG-RRRRR-G-RRRRR amide was reacted with NHS-mPEG (*O*-[(*N*-succinimidyl)succinyl-amino ethyl]-*O'*-methyl-poly(ethylene glycol) at the end of the cleavage domain; the italicized amino acids correspond to the MMP-9 substrate. The peptide consists of a cleavage domain with the enzyme recognition motif and a highly
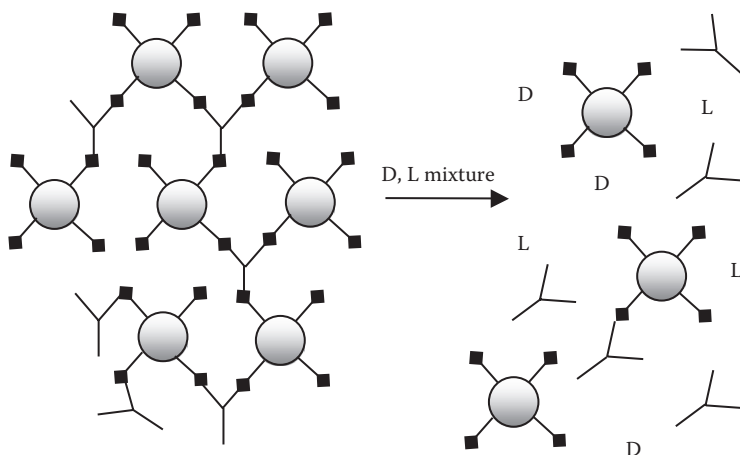
FIGURE 7.11
The addition of d-Phe impurities contained in l-Phe samples causes dispersion of CLIO–d-Phe/anti-d-AA self-assemblies and a corresponding increase in the T_2 relaxation time. (After Schellenberger, E., et al., *Bioconjugate Chem.*, 19, 2440, 2008.)

positively charged, arginine ($C_6H_{14}N_4O_2$)-rich coupling domain, which are connected by a linker sequence. For analysis by fluorescence methods, a fluorescein is coupled to the peptide. The resulting MMP-9-peptide-mPEG copolymer was purified by gel filtration; mPEG = $CH_3(OCH_2CH_2)_nOH$. To prepare 6×-MMP-9-PSOP (protease-specific iron oxide particles), with six peptide-mPEG copolymers per VSOP, in HEPES buffer (pH = 7.5) were mixed with peptide-mPEG in HEPES [2-[4-(2-hydroxyethyl)piperazin-1-yl] ethanesulfonic acid: $C_8H_{18}N_2O_4S$] buffer and stirred immediately. A ratio of 6 peptide-mPEG per VSOP yielded MMP-9-PSOP with a hydrodynamic diameter of 24.9 ± 7.0 nm. Mixtures with ratios between 6 and 16 peptide-mPEG consistently yielded particles with sizes around 24 nm, whereas ratios below 6 resulted in particles over 30 nm in size.

When the sterically stabilized MMP-9-PSOP were exposed to MMP-9, the protease cleaved the peptide at the recognition site, resulting in loss of the sterically stabilizing mPEG shell. Due to the superparamagnetic properties of the iron oxide cores and the ambivalent surface of the remaining particles with positively (coupling domains) and negatively (citrate ($C_6H_8O_7$) coat of VSOP) charged areas, the particles aggregated driven by magnetic and electrostatic attraction.

This process has two important consequences: First, clustering of the superparamagnetic nanoparticles causes a substantial increase in R_2-relaxivity, called magnetic switch. Second, the particles are converted from mPEG-covered stealth particles, avoiding detection, into highly aggregative particles with strongly charged surfaces. Conveniently, mPEG-5000 has been shown to be optimal for achieving stealth properties for nanoparticles. Consequently,

once injected, the intact PSOP should remain for a long time in the blood circulation until they reach an MMP-9 expressing target tissue, where they are converted into aggregative particles and accumulate.

7.7 Magnetic Relaxation Switch Immunosensors

Tsourkas et al. (2004) developed a homogeneous enantioselective immunosensor that utilizes magnetic relaxation switching. An enantiomer is one of two stereoisomers that are mirror images of each other. "Enantioselective" relates to a chemical reaction in which one enantiomer of a chiral product (a molecule that is not superimposable on its mirror image) is preferentially produced. The enantioselective MRS immunosensor was based on magnetic nanoparticles labeled with a derivative of *d*-phenylalanine (d-Phe), $C_9H_{11}NO_2$, a form of the essential amino acid, *l*-Phenylalanine, that, when taken as a supplement, protects the body's production of naturally occurring painkillers.

The magnetic nanoparticles consisted of a superparamagnetic iron oxide core with an aminated cross-linked dextran coating (CLIO). The CLIO–d-Phe nanoparticles were generated by first coupling tyramine ($C_8H_{11}NO$) to the primary amino groups of the CLIO nanoparticles by using a homobifunctional NHS ester ($C_4H_5NO_3$). Diazotization (reaction between a primary aromatic amine and nitrous acid to give a diazonium compound) was then used to couple p-amino-d-phenylalanine ($C_9H_{12}N_2O_2$) through its side chain to the tyraminyl residues, thus preserving both the a-amino and carboxy groups attached to the stereogenic center (also known as a chiral center, it is characterized by an atom which has different groups bound to it in such a manner that its mirror image is nonsuperimposable).

When antibodies specific to d-amino acids (anti-d-AA) were added to the CLIO–d-Phe nanoparticles (Figure 7.12), the divalent nature of the antibodies resulted in the self-assembly of the nanoparticles, which led to a decrease of more than 100 ms in the T_2 relaxation time. The presence of d-Phe impurities in samples of *l*-Phe was then determined by performing a one-step competitive immunoassay. Upon addition of mixtures of the enantiomers (one of two stereoisomers that are mirror images of each other that are non-superposable) to the CLIO–d-Phe/anti-d-AA self-assembled structures, the presence of d-Phe impurities resulted in the dispersion of the nanoparticles by competing with the CLIO–d-Phe conjugates for antibody binding sites. This subsequently led to an increase in the T_2 relaxation time. The presence of free d-Phe impurities could be detected within minutes, and the rate and magnitude of change in the T_2 relaxation time was dependent on the concentration of impurities.

An important attribute of magnetic relaxation switch immunosensors is their potential for the rapid determination of enantiomeric excess in a high-throughput format.

7.8 Magneto Nanosensor Microarray Biochip

7.8.1 Rationale and Motivation

Medical decision-making is increasingly based on molecular testing. Some of the crucial issues involved are as follows: (i) Detection across varied samples is difficult. A urologist (a physician who has specialized knowledge and skill regarding problems of the male and female urinary tract and the male reproductive organs) provides urine sample, a neurologist
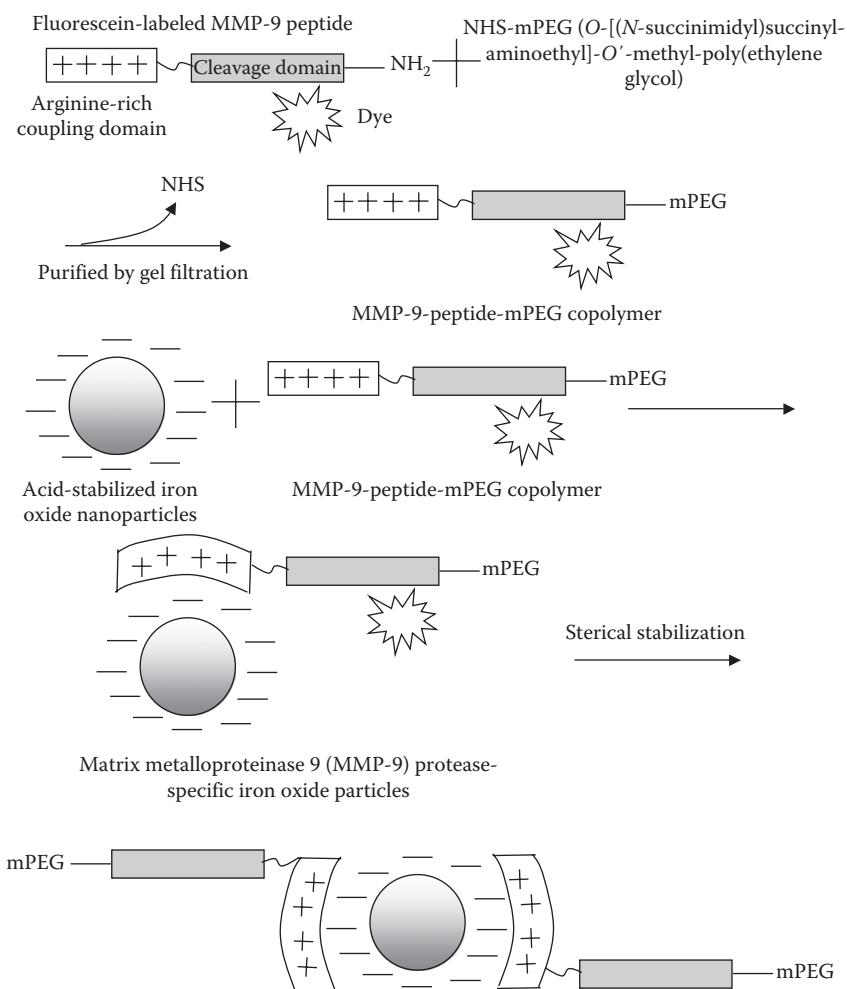
FIGURE 7.12
Synthesis and function of MMP-9-activatable protease-specific iron oxide particles. (After Tsourkas, A. et al., *Angew. Chem.*, 116, 2449, doi: 10.1002/ange.200352998, 2004.)
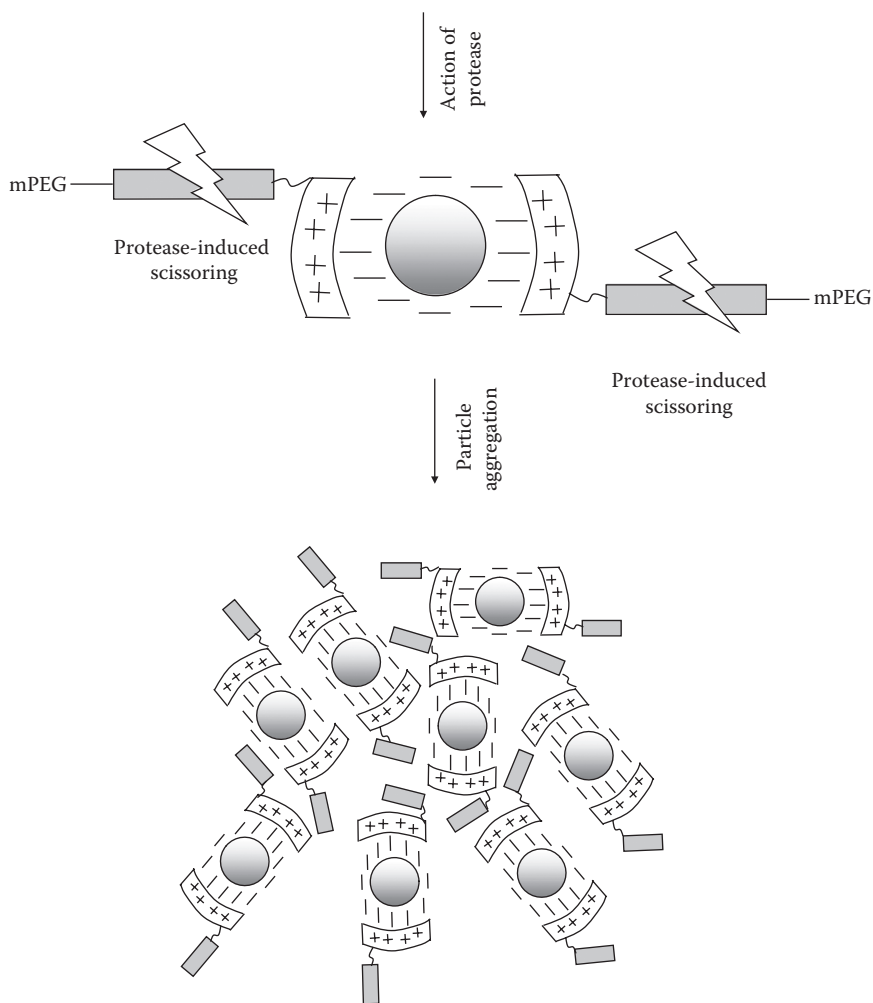
FIGURE 7.12 (Continued)

dealing with disorders of the nervous system gives cerebrospinal fluid (a protective nourishing fluid that circulates around the brain and spinal cord of the central nervous system), and a cardiologist (a heart specialist) provides blood or an oncologist (a physician who studies, diagnoses and treats cancerous tumors) cell lysates (the cellular debris and fluid produced by lysis). The diversity of such supporting matrices has stalled the generalization and sensitivity of protein detection platforms, thus greatly degrading their clinical utility. (ii) Another serious impediment is signal distortion that occurs in various matrices due to heterogeneity in ionic strength, pH, temperature, and autofluorescence (the fluorescence of other substances than the fluorophore of interest increasing the background

signal). (iii) Sample preparation is in itself an elaborate and intricate preparatory step for any analysis. Samples need to be made in pure water or accurately controlled salt solutions

What is the solution to the problem? A matrix-independent method is necessary. A general sensing platform that can be ubiquitously applied to detect the constellation of biomolecules in varied clinical samples (e.g., serum, urine, cell lysates, or saliva) with high sensitivity and large linear dynamic range is the answer to the problem. The additional advantage is that the matrices of even the most complex biological samples lack a detectable magnetic background signal and produce no interference with the magnetic transduction mechanism. Therefore, a magnetic field–based detection platform is most appropriate for protein detection in clinical samples.

Reasoning on the above lines, Gaster et al. (2009) presented a magnetic nanosensing protein detection technology that overcomes the problems associated with other methodologies. They described a magnetic array biochip that is used for detecting surface binding reactions of biological molecules previously labeled with 10–100 nm superparamagnetic nanoparticles.

Is magnetic array biochip similar to DNA fluorescence-based biochip? Yes, in several respects, but magnetic labeling offers distinct advantages as compared to optical fluorescent labeling. *What are these advantages?* Some of the advantages are as follows: (i) Inexpensive chip reader instrumentation, amenable to miniaturization. (ii) Absence of label bleaching. (iii) Improved background rejection. (iv) Potentially higher sensitivity. (v) Ability of simultaneously measuring multiple binding reactions in homogeneous assays requiring no separation steps. Pipetting, incubating, and measuring are the only steps required. (vi) Affording seamless integration with magnetic separation methodologies. (vii) Although biological molecules can exhibit autofluorescence, they have no intrinsic magnetic signal. (viii) Magnetically labeled biomolecules can be manipulated by applied magnetic fields to pre-concentrate the analyte using microfluidic devices.

The basic elements of magnetic array biochip are the *giant magnetoresistive* (*GMR*) *sensors* whose operation is explained in the next section.

7.8.2 Sensor Choice, Design Considerations, Passivation, and Magnetic Nanotag Issues

As already described and for obvious reasons, GMR sensors are favored (Kasatkin et al. 2010). These sensors, originally developed for use as read heads in hard-disk drives, are multilayer thin-film structures. They work on the basis of a quantum-mechanical effect, wherein a change in the strength and orientation of the local magnetic field induces a modification in resistance of the sensor.

The magnetoresistive thin film is sputtered on an insulating substrate silicon wafer and patterned by lithographic techniques to delineate the array of several sensors. The aforementioned magnetic array biochip contained 64

such sensors in an area of 1.2 cm × 1.2 cm. Using a reasonably simple readout circuit, each sensor constituting the array can be independently and simultaneously accessed and signal obtained.

The magnetoresistive sensor responds very fast on the timescale of nanoseconds. *How does the response of magnetoresistive sensors compare with conventional inductive coils?* Response of magnetoresistive sensors is a direct function of the proximate magnetic field whereas the inductive pickup coils respond only to changing magnetic fields.

What are the design considerations for the sensor or biochips? (i) The width of the sensor is an important performance-determining parameter. When very small magnetic nanoparticles are used, the narrower the sensor, the higher the sensitivity. But the benefit thus derived is limited because according to theory, the edge of the spin valve produces an inverse (negative) signal decreasing the positive signal from the reducing area of spin valve. (ii) Magnetic domain formation of the sensor must be paid attention. The manner in which the free layer responds to the changing tickling field is governed by the layout of the sensor. Edges produce a local demagnetizing field favoring the alignment of the magnetization parallel to any edges of the sensor. This can be utilized to instill a bias into the magnetization of the free layer, which will tend to align with the long axis of a linear segment. In the same way, curved or bent segments may inhibit coherent domain rotation degrading the linearity and reproducibility of domain rotations and thereby increasing the noise in the measurements. (iii) The final resistance value obtained from the design is a vital parameter. It should lie in a range that is easily measurable at low voltages. Higher measuring voltages mean applying more stress on the very thin passivation films isolating the sensor from the aqueous reagents.

What are the challenges facing the passivation of the magnetic arrays? It should be sufficiently durable to withstand corrosion in aqueous environment. Leakage currents should be minimized. But it must be thin enough so that sensitivity of the chip is not decreased. An oxide–nitride–oxide (ONO) trilayer of 50 nm thickness serves as a good passivation. The symmetrically placed silicon dioxide layers on the two sides of this composite layer do not allow any residual stress to be developed which is likely in a two-layer structure. Also, silicon nitride is a good diffusion barrier. Many spin valve sensors show no corrosion damage as evidenced by a reproducible signal baseline in repeated measurements and stability of resistance values when immersed in aqueous solutions or buffers. Although corrosion damage may be insignificant but alternating current leakage occurs from parasitic capacitance across the thin passivation layer. This parasitic capacitance is very small so that cross talk among various sensors of an array is not troublesome. Nonetheless, it leads to a "water signal." Consequently, there is a small reversible baseline shift whenever the sensors are transported between dry and wet states. Such kind of baseline shifts disturbances are found in open-well chips in which there are brief dry periods during reagent changes, but are not likely in reagent

recycling in microfluidic chips because they do not require dry spells for complete reagent changes.

What are the important issues concerning magnetic nanoparticles used for tagging? As already indicated, superparamagnetic particles are used to avoid clustering and precipitation. The smaller the nanoparticle, the higher is its rate of diffusion. Further, a small particle helps in restricting the observation volume to surface-bound labels only. Obviously, the nanoparticles should have as high a magnetic moment as possible. But the most critical property of the magnetic nanotag is its surface chemistry and stability in aqueous suspension. Surface chemistry must be carefully controlled to enable highly selective and strong binding reactions. This implies that only molecules of interest are labeled. There should not be any particle precipitation because it continuously raises the signal baseline obscuring the equilibration of binding reactions. This is of special concern at low analyte concentrations.

7.8.3 Understanding Magnetic Array Operation

Let us perform an easy thought experiment in which a miniscule permanent magnet called a magnetic nanotag is attached as a tag or label to a biomolecule of interest (Osterfeld and Wang 2009). Suppose now this biomolecule attaches through a specific binding reaction to a sensor of the magnetic array. Consequent to this attachment, a small change in resistance of the particular sensor of the array will be recorded. For a properly oriented magnetic label, the strength of the signal induced in the magnetoresistive sensor will follow the strength of the dipolar field from the miniscule permanent magnet label; it will be inversely proportional to the cube of the separation distance between the magnetic label and the sensor to which it attaches. In practice, there is an optimal label-to-sensor distance determined by the stray field from the sensor and the curvature of magnetic field lines, which on approach decreases the in-plane component to which the sensor responds. Due to the finite observation volume, properly designed sensors are ideal for detection of surface bound magnetic nanotags.

What is the effect of this finite observation volume? Unbound magnetic labels, if they are sufficiently stable in high concentration in the suspension, are unlikely to penetrate into this observation volume (Figure 7.13) in large numbers so that the experiments are not disturbed by background noise from these unbound labels. Due to the high rejection of unbound labels, removal of the excess unbound labels is not required enabling the execution of homogeneous assays omitting the washing step.

Will it be possible to use permanent magnetic labels in a reality? Obviously not, because the labels will tend to cluster together and precipitate along with the molecules to which they are attached rendering the analysis unfeasible. Their free orientation will result in a small net signal. *How is this situation overcome?* The clustering situation is avoided by using superparamagnetic nanoparticles as labels. The important feature of these particles is that they
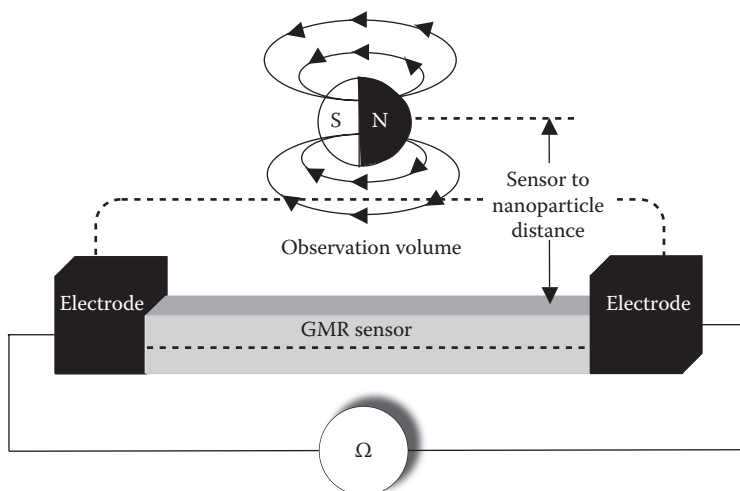
FIGURE 7.13
Observation volume: the signal from a magnetic nanoparticle decreases perceptibly as the sensor to nanoparticle distance increases to a few hundred nanometers resulting in a finite observation volume covering mainly surface-bound nanoparticles. The background signal from the unbound particles is almost negligible. Rejection of this background signal avoids interferences from it. (After Osterfeld, S. J. and Wang, S. X, MagArray biochips for protein and DNA detection with magnetic nanotags: Design, experiment and signal-to-noise ratio, in *Microarrays: Preparation, Microfluidics, Detection Methods and Biological Applications*, Dill, K., R. H. Liu, and P. Grodzinski (Eds.), Springer, New York, 299, 2009.)

have no remnant magnetic moment. For producing a magnetic signal from a collection of superparamagnetic nanolabels, a directed magnetic ticking field is applied, driving the sensor array from its equilibrium position. The tickling field also affects the supermagnetic labels because they become magnetized by this field and produce a small dipolar magnetic field opposite in direction to the tickling field itself. This dipolar field of magnetic labels slightly decreases the response of the sensor to the tickling field.

The magnetic array strategy employs a "sandwich" assay in which the target antigen is sandwiched between two antibodies, one bound to the sensor and the other tagged with a superparamagnetic nanoparticle. Under an external magnetic field, the nanoparticles magnetize, and their presence or absence can be detected by the underlying GMR sensor.

What is the general methodology for conducting an assay? The main steps of the assay are as follows: (i) Capture antibodies that are complementary to a chosen antigen of curiosity are immobilized onto the surface of each sensor. (ii) The complementary antigens bind to the antibodies while the noncomplementary antigens are washed off. (iii) After adding a concoction of detection antibodies, the biotinylated detection antibody complementary to the antigen of interest fastens in a sandwich structure, and the noncomplementary antibodies are washed away. (iv) Finally, a streptavidin-labeled magnetic

nanoparticle tag is added to the solution, and it binds the biotinylated detection antibody; biotin $= C_{10}H_{16}N_2O_3S$. (v) As the magnetic tags diffuse to the GMR sensor surface and bind the detection antibody, the magnetic fields from the magnetic nanoparticles are detected by the underlying MR sensor in real time in the presence of a small external modulation magnetic field.

Osterfeld and Wang (2009) described a systematic procedure to carry out a high-sensitivity proteomics assay for quantifying the concentration of biotinylated anti IFN-γ in phosphate buffered saline (PBS) buffer solution. Magnetic array is shown in Figure 7.14a. Sensor A has a thick passivation; therefore, it remains unaffected by magnetic nanoparticles. Because nanoparticles are prevented from entering the physical barrier of observation volume for this sensor A by the passivation, clearly, sensor A will not respond to any magnetic nanoparticle label. It will maintain an absolute reference level and serves as a *reference sensor* for the measurements. Sensor B is coated by IFN-γ in PBS buffer manually by pouring a droplet over it. Such a sensor is called a *positive sensor* and is expected to interact with the analyte to be examined. Sensors like sensor C are not covered with any biomolecule and are therefore susceptible to nonspecific adsorption of IFN-γ or any molecules passing by. Hence, such sensors are termed *neutral sensors*. Sensors like sensor D are saturated with IL6-sR and will therefore not be vulnerable to any nonspecific adsorption like sensor C. They are known as *negative sensors*. Using the procedure outlined above, portions of the magnetic array were functionalized in one of the four ways to produce reference sensors (A), positive sensors (B), neutral sensors (C), and negative sensors (D), as shown in Figure 7.14b.

After functionalizing the chip, it is rinsed with 1% BSA in PBS buffer solution and the chip well is filled with analyte containing anti-IFN-γ in PBS buffer. The chip is rinsed again with BSA and transferred to measuring station for analyte quantification. Two pairs of differential signals are recorded: (i) One pair is between positive sensor and reference sensor indicating the amount of specific adsorption. (ii) Another pair is between negative and reference sensors revealing the amount of nonspecific adsorption on a non-matching functionalization, hence giving the cross-reactivity. An additional differential pair between signals on neutral and reference sensors will indicate the amount of nonspecific adsorption on a bare surface.

After the differential pairs have been established as desired, the chip is subjected to priming rinses in PBS buffer several times. During the alternating wet and dry transitions between these priming rinses, the baseline shifts back and forth but these shifts are reversible and reproducible. After stabilization, the streptavidin-coated nanoparticle solution is delivered to the reaction well of the chip and incubated. Capturing of these nanoparticles by the biotin ligand of the analyte causes their immobilization within the observation volume of the sensors, accompanied by a rise in the signal level. After incubation period, the excess nanoparticles are flushed out of the reaction well by air followed by rinsing with deionized water.
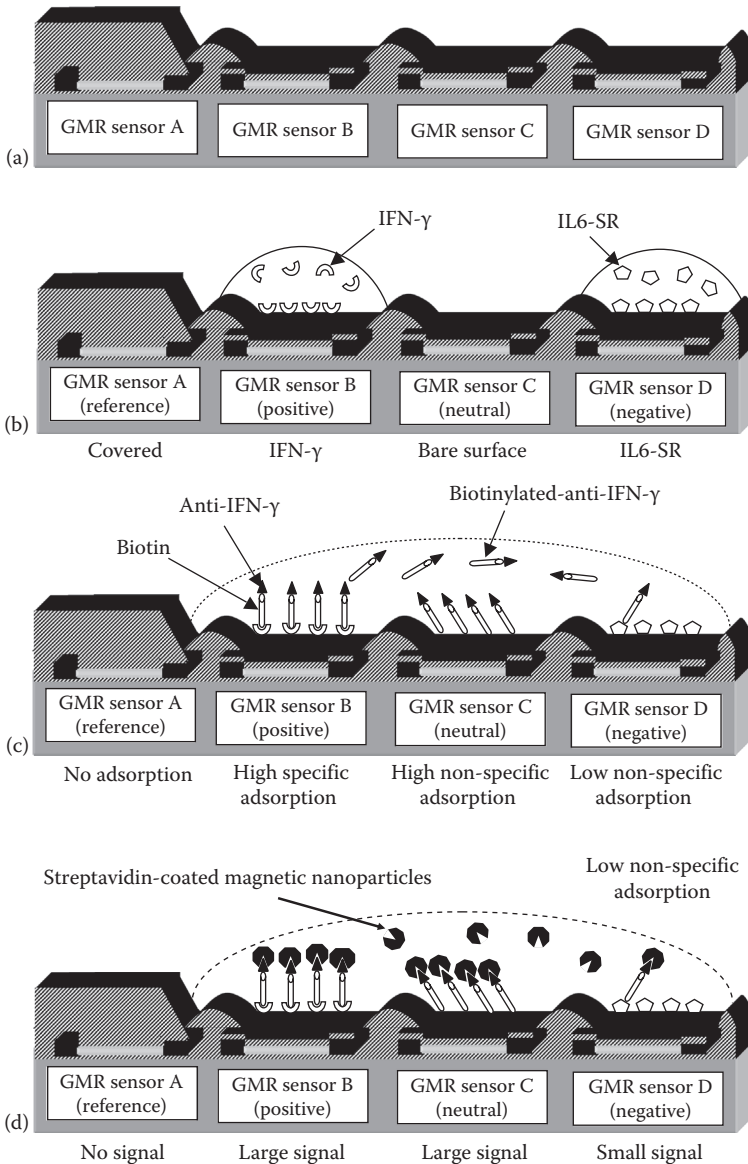
FIGURE 7.14

Procedure of conducting an assay for biotinylated anti IFN-γ: (a) magnetic array, (b) functionalization of chosen sensors with IFN-γ and IL6-sR by droplet coating, (c) incubation of the analyte, and (d) real-time magnetic measurement when magnetic nanoparticle labels are grabbed by analyte molecules and binding of the particles takes place. (After Osterfeld, S. J. and Wang, S. X, MagArray biochips for protein and DNA detection with magnetic nanotags: Design, experiment and signal-to-noise ratio, in *Microarrays: Preparation, Microfluidics, Detection Methods and Biological Applications*, Dill, K., R. H. Liu, and P. Grodzinski (Eds.), Springer, New York, 299, 2009.)

An important figure of merit of the array is the *biochemical signal-to-noise ratio* defined as the ratio of positive and negative signals taking into account nonspecific adsorption and cross-reactivity.

7.8.4 Influence of Reaction Conditions on the Sensor

Gaster et al. (2009) investigated how the sensor itself (before addition of the detection antibody) responded to various reaction conditions, including pH, temperature, and turbidity (the cloudiness or haziness of a fluid). In contrast to nanowires, in which a change of 0.5 pH causes considerable signal fluctuations, their sensing technology was unaffected by changes in ionic strength and pH change between pH 4 and 10. In addition, unlike microcantilevers, for which even a 0.5°C change causes substantial cantilever deflection, their sensors were unaffected by changes in the temperature of sample, provided that a simple temperature correction algorithm was implemented. This is performed in real time without having to rely on reference sensors. Finally, optical activity or turbidity of sample solutions had no effect on this detection platform, as it does not use optical-based detection methods as do enzyme-linked immunosorbent assays (ELISAs), protein microarrays, and quantum dots. Protein arrays comprise a library of proteins or antibodies immobilized in an ordered manner in a 2D addressable grid on a chip. They are used to identify protein–protein interactions, to identify the substrates of protein kinases, or to identify the targets of biologically active small molecules.

7.8.5 DNA and Tumor Marker Detection

A DNA microarray (also commonly known as gene chip, DNA chip, or biochip) is a collection of microscopic DNA spots attached to a solid surface. It is used to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome.

Xu et al. (2008) developed a GMR biochip based on spin valve sensor array and magnetic nanoparticle labels for inexpensive, sensitive, and reliable DNA detection. The resistance of a GMR sensor changes with the magnetic field applied to the sensor, so a magnetically labeled biomolecule can induce a signal. Human papillomavirus (HPV) is a subset of papillomaviruses that infects the epithelial cells of the skin and mucus membranes in humans. Infection with HPV is associated with various forms of cancers, including cervical cancer, cancer that starts in the cervix, the lower part of the uterus (womb) that opens at the top of the vagina.

The GMR sensor had a bottom spin valve structure: Si/Ta(5)/seed ayer/ IrMn(8)/CoFe(2)/Ru/(0.8)/CoFe(2)/Cu(2.3)/CoFe(1.5)/Ta(3), all numbers in parenthesis are in nanometers. Each chip has 32 pairs of GMR sensors, which are connected to the bonding pads on the periphery by a 300 nm thick

Ta/Au/Ta lead. Each sensor consists of 32 spin valve strips in serial connection. Each strip has an electrical active area of $93\,\mu m \times 1.5\,\mu m$. To protect the sensors and leads from corrosion, two passivation layers were deposited by ion beam sputtering: first, a thin passivation layer of $SiO_2(10\,nm)/Si_3N_4(20\,nm)/SiO_2(10\,nm)$ was deposited above all sensors and leads, exposing only the bonding pad area; second a thick passivation layer of $SiO_2(100\,nm)/Si_3N_4(150\,nm)/SiO_2(100\,nm)$ was deposited on top of the reference sensors and leads, exposing the active sensors and bonding pad area. In the absence of an applied field, the total resistance of one sensor is about $35\,k\Omega$. The sensors showed an accuracy of $\sim90\%$, with good signal consistency across chips.

Gaster et al. (2009) applied magneto-nanosensors to simultaneously monitor real-time binding events of multiple tumor markers in a number of biological fluids. Their technology employs a "sandwich" assay in which the target antigen is sandwiched between two antibodies, one bound to the sensor and the other tagged with a superparamagnetic nanoparticle. The basic mechanism is to capture antigens—deleterious compounds produced and shed by the cancer cells—using antibodies that naturally tend to bond with the antigens. The antibodies, termed "capture antibodies," are applied to a sensor, so that when the matrix of interest is placed onto the sensor chip, the appropriate antigens bind. While the antigens are held fast, another dollop or lump of the antibodies is applied. These antibodies are attracted to the antigens held on the sensors, and in bonding with them effectively seal the antigens inside an antibody sandwich. The researchers then apply a wash containing magnetic nanoparticle tags that have been tailored to fit specific antibodies. The magnetic nanotags attach themselves to the outer antibody on the sandwich. Under an external magnetic field, the nanoparticles magnetize, where they alter the ambient magnetic field in a small but distinct and detectable way, and their presence or absence can be detected by the underlying GMR sensor (Figure 7.15).

They functionalized magnetosensors with antibodies to a representative panel of tumor markers. They were also able to perform protein detection in human serum, human urine, and human saliva (also known as spit, is a clear liquid that's made in mouth 24 h a day, every day). Using chips measuring $1.2\,cm \times 1\,cm$, each containing an array of 64 GMR individually addressable magneto-nanosensors, they showed real-time measurements of protein concentrations down to the attomolar level in a variety of clinically relevant media with a linear dynamic range (that concentration over which the sensor output is linearly related to the analyte concentration) of over six orders of magnitude. The matrix insensitivity of the platform to various media demonstrates that magnetic nanosensor technology can be directly applied to a variety of settings such as molecular biology, clinical diagnostics, and biodefence.

7.8.6 GMR-Based Detection System with Zeptomole (10^{-21} mol) Sensitivity

Srinivasan et al. (2009) reported the development of a highly sensitive detection system based on a GMR sensor and 12.8 nm high moment cubic FeCo nanoparticles, which linearly detects 600–4500 copies of streptavidin based on the biotin–streptavidin interaction (Figure 7.16). They also demonstrated the feasibility of this detecting system for real biological applications, with the example of the linear detection of human interleukin 6 (IL-6, a potential lung cancer biomarker) through a sandwich-based principle. IL-6 is one of the most important mediators of fever.

The GMR sensor consists of a multilayer structure: Ta(5 nm)/ Ir0.8 Mn0.2(10 nm)/Co0.9Fe0.1(2.5 nm)/Cu(3.3 nm)/Co0.9Fe0.1(1 nm)/ Ni0.82Fe0.12(2 nm)/Ta(5 nm). The GMR sensor surface was sequentially
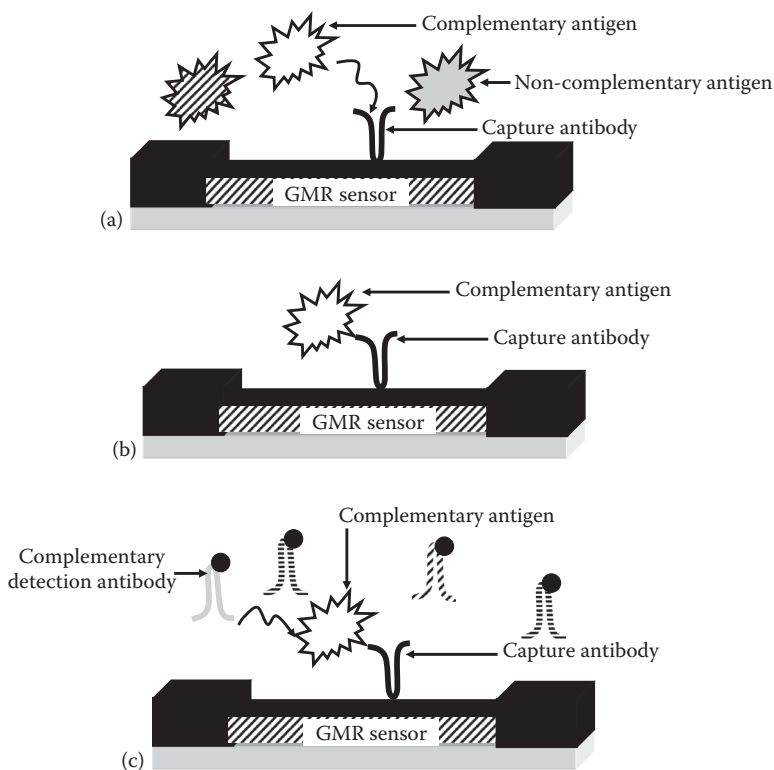
FIGURE 7.15
Sandwich assay: (a) Capture antibodies that are complementary to a chosen antigen are immobilized onto the surface of each sensor. (b) The noncomplementary antigens are subsequently washed away. (c) After adding a cocktail of detection antibodies, the biotinylated detection antibody complementary to the antigen of interest binds in a sandwich structure, and the noncomplementary antibodies are washed away.
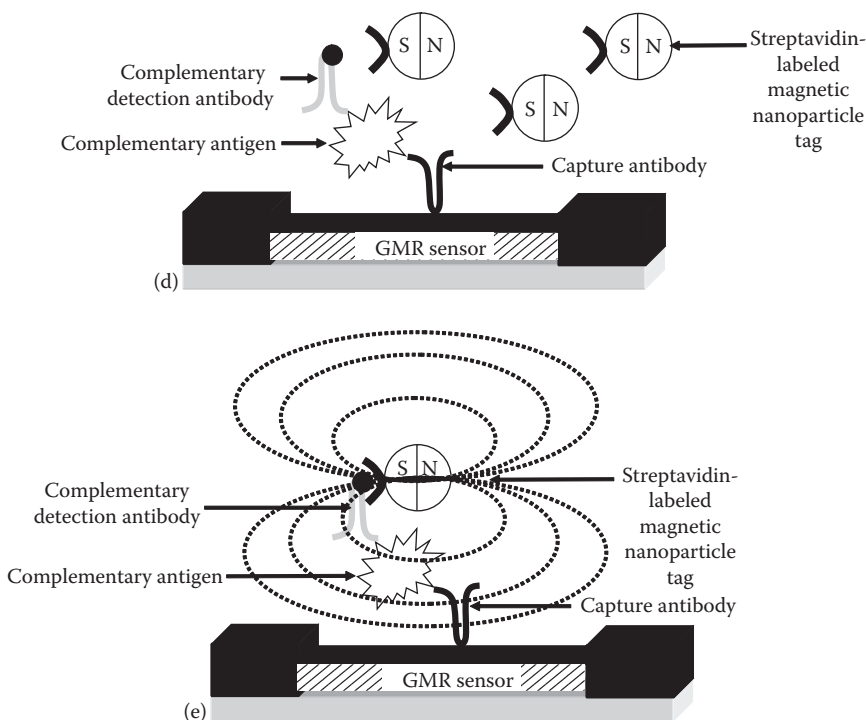
FIGURE 7.15 (Continued)
(d) A streptavidin-labeled magnetic nanoparticle tag is added to the solution, and it binds the biotinylated detection antibody. (e) As the magnetic tags diffuse to the GMR sensor surface and bind with the detection antibody, the magnetic fields from the magnetic nanoparticles can be detected by the underlying GMR sensor in real time in the presence of a small external modulation magnetic field. (After Gaster, R. S. et al., *Nat. Med.*, 15, 1327, 2009.)

modified with 3-aminopropyltriethoxysilane (APTES), $C_9H_{23}NO_3Si$, followed by Chromalink biotin ($C_{38}H_{49}N_8NaO_{13}S_2$) *N*-hydroxysuccinimidyl ester ($C_4H_5NO_3$). High-magnetic-moment FeCo nanoparticles with a composition of 70:30 were synthesized using a sputtering gas condensation technique. In gas condensation, a metallic or inorganic material is vaporized using thermal evaporation sources such as a Joule heated refractory crucibles. An alternative method employs sputtering or laser evaporation. These methods may be used instead of thermal evaporation. Sputtering is a non-thermal process in which surface atoms are physically ejected from the surface by momentum transfer from an energetic bombarding species of atomic/molecular size.

A high residual gas pressure causes the formation of ultrafine particles (100 nm) by gas phase collision. The ultrafine particles are produced by collision of evaporated atoms with residual gas molecules. Gas pressures >3 mPa (10 torr) are required.
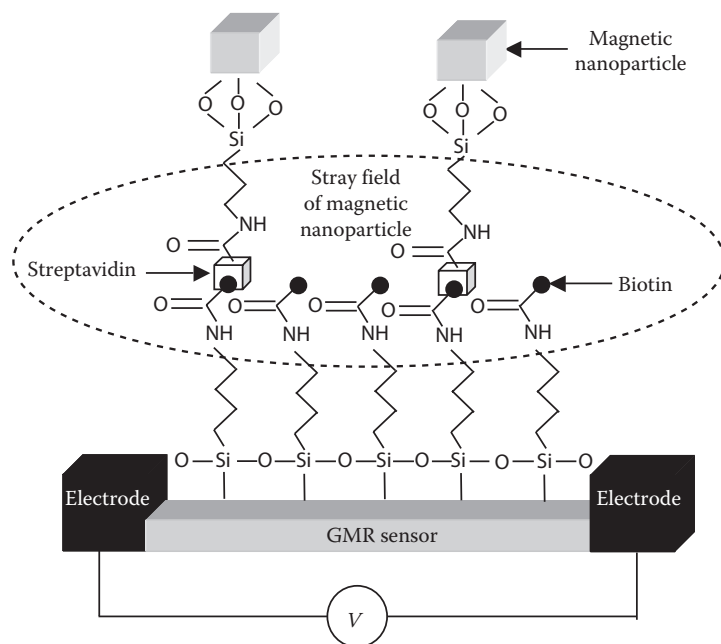
FIGURE 7.16

GMR nanosensor for detecting biotin–streptavidin interaction. (After Srinivasan, B. et al., *Angew. Chem. Int. Ed.*, 48, 2764, 2009.)

The FeCo nanoparticles have an oxidation layer 1.5 nm thick. FeCo nanoparticles were first modified with amino groups on the surface by using APTES; this resulted in approximately 660 copies of APTES molecules on each nanoparticle. APTES-modified nanoparticles were subjected to streptavidin-AF (Alexa Fluor) 488 modification using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC), $C_8H_{17}N_3$ coupling chemistry. EDC is a water-soluble carbodiimide usually obtained as the hydrochloride and generally employed as a carboxyl activating agent for the coupling of primary amines to yield amide bonds. It is estimated that each nanoparticle is modified by 1.3 streptavidin AF 488 molecules. To explore the sensitivity of this detecting system based on GMR sensors and magnetic nanoparticles, varied quantities of streptavidin-AF488-modified magnetic nanoparticles were applied onto the surface of a GMR sensor modified by Chromalink biotin. After thorough washing to remove the potentially unbound streptavidin, magnetic signals from the magnetic nanoparticles specifically retained on the GMR sensor through biotin–streptavidin interactions were measured by the GMR sensor.

The GMR sensor detected signals from as few as 600 copies ($<10^{-21}$ mol, zeptomol) of streptavidin; this sensitivity is expected to suffice for detection of all known potential biomarkers from body fluid samples of 10 nL

or less. More importantly, there is a linear dose–response relationship for the amount of streptavidin applied and the magnetic signals detected by the GMR sensors (saturation of signal was observed with more than 20,000 copies of streptavidin). Such a dynamic range of linearity outperforms most other GMR-based detecting systems reported to date.

7.8.7 Bead ARray Counter (BARC) Biosensor

Microbeads are uniform-size particles, typically 0.5–500 µm in diameter. Bio-reactive molecules can be adsorbed or coupled to their surface, and used to separate biological materials such as cells, proteins, or nucleic acids. Baselt et al. (1998) conducted a proof-of-concept biosensor experiment showing the worth of using GMR sensors with magnetic microbeads in a Bead ARray Counter (BARC). An array of 80 µm × 5 µm GMR sensor elements was fabri-cated from sandwich GMR material. Each sensor was coated with different biological molecules that will bond to different materials to be assayed. The magnetic microbeads were also coated with the materials to be analyzed. The microbeads in suspension were allowed to settle onto the GMR sensor array where specific beads will bond to specific sensors only if the materials are designed to attract each other. Non-binding beads were removed by a small magnetic field. The beads were then magnetized at 200 Hz by an ac electro-magnet. The 1 µm microbeads were made up of nm sized iron oxide particles which had little or no magnetization in the absence of an applied field. A lock-in amplifier extracted the signal at twice the exciting frequency from a Wheatstone bridge constructed of two GMR sensor elements, one of which was used as a reference, and two normal resistors. High-pass filters were used to eliminate offset and the necessity of balancing the two GMR sensor ele-ments. With this detection system, the presence of as few as one microbead was detectable.

7.9 Needle-Type SV-GMR Sensor for Biomedical Applications

Cancer is the most deadly disease in the world today. Hyperthermia is medi-cal treatment method for this disease by the induction of fever, as by the injection of a foreign protein or the application of heat. It is a cancer treat-ment method that utilizes the property that cancer cells are more sensitive to temperature than normal cells. This method is used for treating cancers of the prostate, liver, bladder, breast, and others. Magnetic hyperthermia is the name given to the treatment based on the fact that magnetic nanopar-ticles, when subjected to an alternating magnetic field, produce heat. Hence, if magnetic nanoparticles are put inside a tumor and the whole patient is

placed in an alternating magnetic field of well-chosen amplitude and frequency, the tumor temperature will rise. This could kill the tumor cells by necrosis (the death of living cells or tissues when there is not enough blood flowing to the tissue, whether from injury, radiation, or chemicals) if the temperature is above 45°C. It can improve the efficiency of chemotherapy if the temperature is raised around 42°C. A hyperthermia cancer treatment based on induction heating is achieved by injecting magnetic fluid with magnetic nanoparticles into the tumor. The control of temperature is an important task in achieving success using this treatment method.

The SVGMR-based needle-type sensor (Mukhopadhyay et al. 2007, Gooneratne et al. 2008) is used to measure the magnetic flux density of the magnetic fluid inside the human body from which the temperature is estimated. The magnetic fluid (magnetite) is usually injected into human body to kill cancerous cell using hyperthermia-based treatment. To control the heat treatment, an accurate knowledge of temperature is essential. The SV-GMR element (Figure 7.17) with sensing area of $75\,\mu m \times 40\,\mu m$ is on the tip of the needle. Sensing direction is parallel to the needle. The volume/weight density of the magnetic fluid is estimated by measuring and comparing the applied magnetic flux density (outside body) and the flux density in the magnetic fluid. A constant current of 0.5 mA is applied to the SV-GMR sensor. The sensitivity is approximately $10\,\mu V\,\mu T^{-1}$.

7.10 Superconductive Magnetic Nanosensor

Granata et al. (2008) fabricated a nano-SQUID (superconducting quantum interference device) for measurement of petite local magnetic signals arising from small atomic or molecular populations. They presented a nano-SQUID with an effective area of $4 \times 10^{-2}\,\mu m^2$; based on nanometric niobium (Nb) Dayem bridge junctions (Figure 7.18). The Dayem bridge is a thin-film variant

(a)   (b)
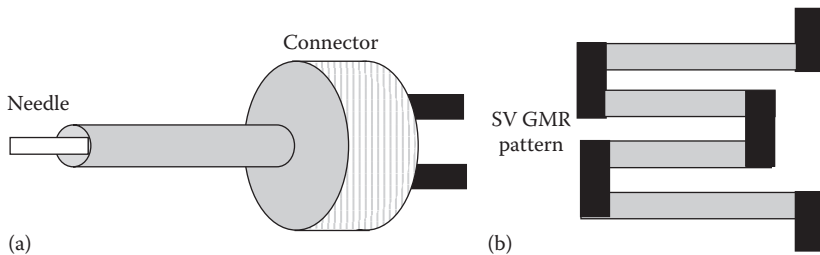
FIGURE 7.17
SV GMR needle sensor: (a) Structure of the probe and (b) spin value giant magneto-resistive pattern. (After Mukhopadhyay, S. C. et al., *IEEE Sens.* J., 7, 401, 2007; Gooneratne, C. P. et al., *J. Sens.*, Article ID 890293, 7, doi: 10.1155/2008/890293, 2008.)
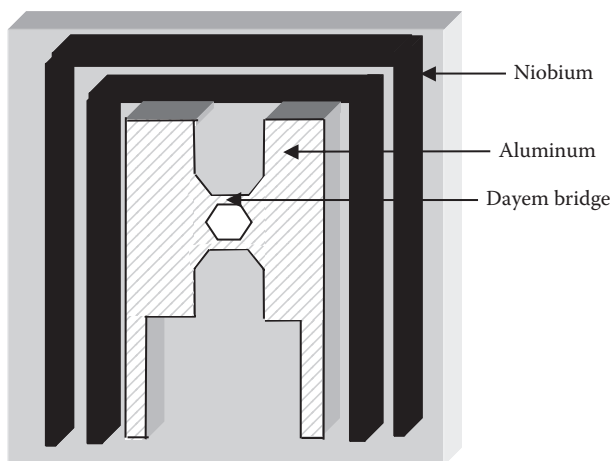
FIGURE 7.18
Nano-SQUID based on niobium Dayem nanobridges. (After Granata, C., et al., *Nanotechnol.*, 19, 275501, doi: 10.1088/0957-4484/19/27/275501, 2008.)

of the Josephson junction, with dimensions on the scale of a few micrometers or less. The sensor, having a washer shape with a hole of 200 nm and two Josephson–Dayem nanobridges of 80 nm × 100 nm, consists of a Nb(30 nm)/Al(30 nm) bilayer patterned by electron beam lithography (EBL) and shaped by lift-off and reactive ion etch (RIE) processes. The presence of the niobium coils, integrated on-chip and tightly coupled to the SQUID, allows to easily excite the sensor in order to get the voltage–flux characteristics and to flux bias the SQUID at its optimal point. The measurements were performed at liquid helium temperature. A voltage swing of 75 μV and a maximum voltage–flux transfer coefficient (responsivity) as high as 1 mV/ϕ_0 were directly measured from the voltage–flux characteristic. The detection of the spin-flips of magnetic atom clusters is a very interesting application of nano-SQUIDs. Possible applications of this nanosensor are envisaged in magnetic detection of nanoparticles and small clusters of atoms and molecules, in the measurement of nanoobject magnetization, and in quantum computing.

7.11 Electron Tunneling–Based Magnetic Field Sensor

Here, the polymer material (PMMA) is used to fabricate inexpensive, batch-fabricated, high-yield, and highly sensitive magnetic field sensors based on tunneling sensor platform (Wang et al. 2006). PMMA or poly(methyl methacrylate), $(C_5O_2H_8)_n$, is a clear plastic, used as a shatterproof replacement for glass. The PMMA magnetic field sensor (Figure 7.19) includes mechanical components and three electrodes. The mechanical components are
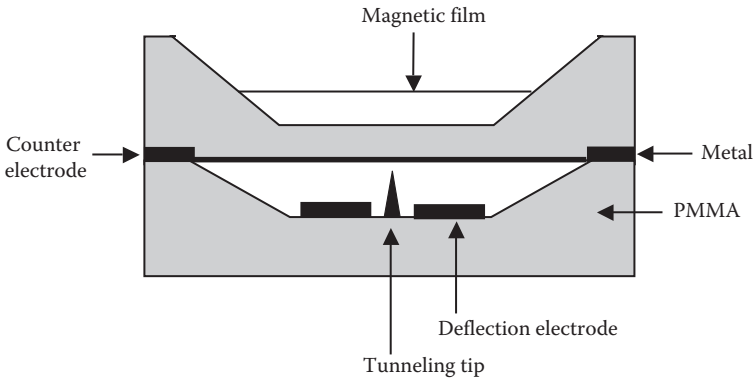
FIGURE 7.19
Micromachined magnetic field tunneling sensor. (After Wang, J. et al., *IEEE Sens. J*, 6, 97, 2006.)

composed of a membrane with a ferromagnetic film at the top surface and a tunneling tip opposite to the film at the bottom. The electrodes include a tip electrode, a counter electrode under ferromagnetic film, and a deflection electrode. The metallic tip electrode and the counter electrode form a tunneling junction. The tunneling current varies exponentially with the gap change.

The titanium/gold (Ti/Au) bilayer is chosen as the electrode metal for its inert chemical characteristics as well as the relatively high work function (the smallest energy in electron volts needed to remove an electron at $0\,K$ to a point immediately outside the solid surface) of gold. When operating at the closed-loop mode, the sensor maintains a constant tip-to-membrane distance by applying an electrostatic feedback force on the deflection electrode.

The displacement changes are caused by the ferromagnetic force when alternating field is applied to the membrane. The transfer function between the output voltage and the ferromagnetic force is derived. The output voltage of the sensor is proportional to the magnetic force applied to the magnetic film. The magnitude of the transfer function is determined by the radius and the thickness of the membrane. The transfer function of a PMMA sensor is substantially large compared with a silicon sensor if the structures remain the same. This is the main reason why polymer is selected as the alternative material instead of silicon.

A natural frequency response of 1.3 kHz was observed, and a tunneling barrier height of 0.713 eV was tested. Due to the quadratic relation (one variable is related to another by a function involving constant terms and terms of first order or higher) between magnetic force and the field, the sensor field response ($7\,0 \times 10^6\,V\,T^{-2}$) was also quadratic. The noise voltage at 1 kHz was 0.2 mV, corresponding to a magnetic field of $0.46 \times 10^6\,T$. The bandwidth of this sensor was 18 kHz.

7.12 Nanowire Magnetic Compass and Position Sensor

Peczalski (2010) reported a nanowire magnetic sensor and position sensor for determining the position of a magnetic object and direction of magnetic field. The magnetic compass includes a plurality of magnetic nanosensors printed on a flexible substrate, which covers 360° angle at equal intervals. Each magnetic nanosensor generally includes magnetoresistive nanowires with high magnetic sensitivity printed in sets of 10 on the flexible substrate. Individual nanosensors are connected into resistive bridge configuration using AMR wires and/or GMR wires. The flexible substrate is also bent to form a circular configuration to detect the azimuth direction (geographic orientation of a line given as an angle measured in degrees clockwise from north) of the magnetic field. The magnetoresistive nanosensors are connected utilizing different bridge configurations for AMR and GMR wires, for example, the Wheatstone bridge connection wherein the magnetic nanosensors are interconnected by metallization. By utilizing multiple nanosensors, the capability of extended angular or linear position measurements is enhanced. An array of magnetic nanosensors is created. The size of the linear array of magnetic nanosensors and the small dimension of the magnetoresistive nanowires enables very high resolution of about 1–0.1 mdeg.

7.13 Discussion and Conclusions

As outlined in the foregoing sections, two main branching points of the magnetic sensor family tree are abundantly clear. On one branch are perched the nanosensors based on GMR principle which obviously has roots in nanotechnology. It can be considered as a major breakthrough in physics, which has impacted our everyday lives profoundly. On the other branch are resident the sensors working as proximity switches and depending on the MRI principle of T_2 variation of adjoining water molecules. Most of the research papers have tended to pursue either of the two directions. Besides, there are developments in nano-SQUID devices and other devices based on quantum-mechanical tunneling principle for magnetic field measurement. Table 7.1 summarizes a few of the gorgeous magnetic nanosensor examples.

Review Exercises

7.1 What is a magnetoresistive effect? Is GMR a magnetoresistive effect? Is a multilayer structure essential for observing GMR? What is GMR ratio?

TABLE 7.1

The Magnetic Nanosensor Family

Sl. No.	Name of the Nanosensor	Operating Principle
1.	Giant magnetoresistance (GMR) sensor	Magnetic-field-dependent resistivity changes due to the quantum-mechanical GMR effect, which arises from the spin-dependent scattering of conduction electrons in ferromagnetic materials. As fabricated, the magnetic moments between adjacent magnetic layers are antiparallel due to antiferromagnetic exchange coupling, and the spin-dependent electron scattering in the multilayer structure is at a maximum. This situation corresponds to the high resistance state of the device. If, however, an external magnetic field is applied in the plane of the multilayer slab, exchange coupling can be overcome and the spins in the magnetic layers brought into alignment with the external field. This alignment lowers the spin-dependent electron scattering and, therefore, the resistance of the GMR. Importantly, the degree of parallel/antiparallel alignment of the ferromagnetic layers is extremely sensitive to the strength of applied magnetic field. As a consequence, GMRs can show up to an 80% change in resistivity at room temperature and at modest magnetic fields (e.g., 10–2000 Oe)
2.	Tunneling magnetoresistance (TMR) sensor	Spin-dependent tunneling (SDT), which results in a change in effective resistance caused by a change in the applied field. The magnitude of the tunneling current between the two magnetic layers is modulated by the direction between the magnetization vectors in the two layers. The resistance versus field effect is similar to the usual GMR spin valve effect but of larger magnitude
3.	Magnetic nanoparticle probes	MR relaxation properties of superparamagnetic nanoparticles such as relaxivity dependence on their degree of clustering. The relaxation induced by these particles is caused by the diffusion of water protons in the inhomogeneous magnetic field surrounding the particles. Magnetic nanoparticles are employed as *proximity sensors* to modulate the spin–spin relaxation time of neighboring water molecules
4.	Protease-specific nanosensor	Upon specific protease cleavage, the nanoparticles rapidly switch from a stable low-relaxivity stealth state to become adhesive, aggregating particles exhibiting increased T_2-relaxivity
5.	Relaxation switch immunosensor	When antibodies specific to D-amino acids (anti-d-AA) are added to the CLIO–d-Phe nanoparticles, the divalent nature of the antibodies results in the self-assembly of the nanoparticles, leading to a decrease of more than 100 ms in the T_2 relaxation time
6.	Superconductive magnetic nanosensor	Nano-SQUID based on nanometric niobium Dayem bridge junctions
7.	Tunneling-based magnetic field sensor	Electron tunneling displacement sensing and relation between magnetic field and magnetic force
9.	Nanowire magnetic compass and position sensor	Magnetoresistive nanowires are utilized as a position sensor of a magnetic object for position determination and finding the direction of magnetic field

7.2 What is the implication of "Giant" in the term "Giant Magnetoresistance"? Does GMR involve any nanoscale phenomenon? What is the nano feature in GMR?

7.3 The AMR effect alters a resistance of a magnetoresistive element in proportion to a square of a cosine of an angle formed between the magnetization of the element and the direction in which a sense current flows through the element. Does the GMR effect depend on the angle between the current flow and magnetic field?

7.4 GMR was independently discovered in 1988 in Fe/Cr/Fe trilayers by a research team led by Peter Grünberg of the Jülich Research Centre (Germany), and in Fe/Cr multilayers by the group of Albert Fert of the University of Paris-Sud (France). What prestigious award was given to them in 2007?

7.5 Is the word "Giant" in GMR not incongruous for a nanotechnology device? Is GMR a quantum-mechanical effect? Explain. How do you distinguish between AMR and GMR?

7.6 While older MR heads typically exhibit a resistance change when passing from one magnetic polarity to another of about 2%, for GMR heads this is anywhere from 5% to 8%. What does this mean in terms of sensitivity, noise, and realizable storage capacity?

7.7 GMR heads are made up of four layers of thin material that combine into a single structure. Name these four layers and describe the function of each layer.

7.8 What is the role of pinned and exchange layers in a spin valve GMR?

7.9 Name the four principal GMR structures and describe their operation with diagrams.

7.10 Discovery and application of the GMR phenomenon is responsible for the ubiquitous availability of economical, high-density information storage devices like compact 160 GB Mp3 players and 1 TB hard drives. Explain.

7.11 What is TMR? Explain SDT phenomenon and the resistance changes in a TMR structure.

7.12 How are magnetic nanoparticles used for detecting specific DNA sequences by structured self-assembly of the disperse particles into a stable assembly (clusters)? Why are they called magnetic nanoswitches?

7.13 Magnetic nanoparticle conjugates that self-assemble in the presence of peroxidases can serve as an *in vivo* diagnostic tool to identify peroxidase-induced diseases by magnetic resonance imaging. Explain the application of peroxidase substrate nanosensors for MRI.

7.14 Give two biosensor examples in which hybridization-induced assembly of magnetic nanoparticles causes a decrease in the spin–spin relaxation time of neighboring water protons.

7.15 What are aminated cross-linked iron oxide nanoparticles (amino-CLIOs)? What is the function of amino groups in these particles?

7.16 What is the advantage of the iron oxide crystal core in CLIO being superparamagnetic? What will happen if it is ferromagnetic?

7.17 How are magnetic nanosensors designed to detect a specific protein in solution by using antibody-mediated interactions? How are they used for measuring various enzymatic activities?

7.18 What are the advantages of a magnetic nanosensor assay as compared to optical assay? What do you mean by "matrix-insensitive assay"? Why is the assay required to be matrix independent?

7.19 Justify the use of a magnetic nanoparticle-based assay as a matrix-insensitive protein detection assay.

7.20 How is a highly sensitive and specific multiplexed detection of protein tumor markers in a matrix-insensitive assay realized by a magnetosensor array?

References

Andreev, S. and P. Dimitrova. 2005. Anisotropic magnetoresistance integrated sensors. *Journal of Optoelectronics and Advanced Materials* 7(1): 199–206.

Arana, S., N. Arana, R. Gracia, and E. Castaño. 2005. High sensitivity linear position sensor developed using granular Ag-Co giant magnetoresistances. *Sensors and Actuators A: Physical* 123: 116–121.

Arana, S., E. Castaño, and F. J. Gracia. 2004. High temperature circular position sensor based on a giant magnetoresistance nanogranular AgxCo1−x alloy. *IEEE Sensor Journal* 4: 221–225.

Baselt, D. R., G. U. Lee, M. Natesan, S. W. Metzger, P. E. Sheehan, and R. J. Colton. 1998. A biosensor based on magnetoresistance technology. *Biosensors & Bioelectronics* 13: 731–739.

Cockburn, B. F. 2004. Tutorial on magnetic tunnel junction magnetoresistive random-access memory. In: *Records of the 2004 International Workshop on Memory Technology, Design and Testing* (MTDT'04), 1087-4852/04, San Jose, CA.

Freitas, P. P., R. Ferreira, S. Cardoso, and F. Cardoso. 2007. Magnetoresistive sensors. *Journal of Physics: Condensed Matter* 19: 165221, doi: 10.1088/0953-8984/19/16/165221

Gaster, R. S., D. A. Hall, C. H. Nielsen, S. J. Osterfeld, H. Yu, K. E. Mach, R. J. Wilson, B. Murmann, J. C. Liao, S. S. Gambhir, and S. X. Wang. 2009. Matix-insensitive protein assays push the limits of biosensors in medicine. *Nature Medicine* 15(11): 1327–1332.

Gooneratne, C. P., S. C. Mukhopadhyay, and S. Yamada. 2008. An SV-GMR needle sensor-based estimation of volume density of magnetic fluid inside human body. *Journal of Sensors* 2008: Article ID 890293, 7pp., doi: 10.1155/2008/890293

Gossuin, Y., P. Gillis, A. Hocq, Q. L. Vuong, and A. Roch. 2009. Magnetic resonance relaxation properties of superparamagnetic particles. *WIREs Nanomedicine and Nanobiotechnology* 1: 299–310.

Granata, C., E. Esposito, A. Vettoliere, L. Petti, and M. Russo. 2008. An integrated superconductive magnetic nanosensor for high-sensitivity nanoscale applications. *Nanotechnology* 19: 275501 (6pp.), doi: 10.1088/0957-4484/19/27/275501

Grimm, J., J. M. Perez, L. Josephson, and R. Weissleder. 2004. Novel nanosensors for rapid analysis of telomerase activity. *Cancer Research* 64: 639–643.

Gupta, A., S. Mohanan, M. Kinyanjui, A. Chuvilin, U. Kaiser, and U. Herr. 2010. Influence of nano-oxide layer on the giant magnetoresistance and exchange bias of NiMn/Co/Cu/Co spin valve sensors. *Journal of Applied Physics* 107: 093910-1–093910-5.

Haun, J., T.-J. Yoon, H. Lee, and R. Weissleder. 2010. Magnetic nanoparticle biosensors. *WIREs Nanomedicine and Nanobiotechnology* 2: 291–304.

Jander, A., C. Smith, and R. Schneider. 2005. Magnetoresistive sensors for nondestructive evaluation. In: *The 10th SPIE International Symposium, Nondestructive Evaluation for Health Monitoring and Diagnostics*, Conference 5770, Copyright © 2005, NVE Corporation, March 6–10, 2005, San Diego, CA, 13pp.

Josephson, L., J. M. Perez, and R. W. Weissleder. 2001. Magnetic nanosensors for the detection of oligonucleotide sequences. *Angewandte Chemie* 2001 113: 3304; *Angewandte Chemie—International Edition* 2001 40: 3204–3206.

Kasatkin, S. I., P. I. Nikitin, A. M. Muravjev, V. V. Lopatin, F. F. Popadinetz, A. V. Svatkov, A. Yu. Toporov, F. A. Pudonin, and A. I. Nikitin. 2000. Spin-tunneling magnetoresistive sensors. *Sensors and Actuators* 85: 221–226.

Kasatkin, S. I., N. P. Vasil'eva, and A. M. Murav'ev. 2010. Biosensors based on the thin-film magnetoresistive sensors. *Automation and Remote Control* 71(1): 156–166.

Kim, G. Y., L. Josephson, R. Langer, and M. J. Cima. 2007. Magnetic relaxation switch detection of human chorionic gonadotrophin. *Bioconjugate Chemistry* 18: 2024–2028.

Koh, I. and L. Josephson. 2009. Magnetic nanoparticle sensors. *Sensors* 9: 8130–8145, doi: 10.3390/s91008130

Millen, R. L., T. Kawaguchi, M. C. Granger, and M. D. Porter. 2005. Giant magnetoresistive sensors and superparamagnetic nanoparticles: A chip-scale detection strategy for immunosorbent assays. *Analytical Chemistry* 77: 6581–6587.

Mukhopadhyay, S. C., K. Chomsuwan, C. P. Gooneratne, and S. Yamada. 2007. A novel needle-type SV-GMR sensor for biomedical applications. *IEEE Sensors Journal* 7(3): 401–408.

Osterfeld, S. J. and S. X. Wang. 2009. MagArray biochips for protein and DNA detection with magnetic nanotags: Design, experiment and signal-to-noise ratio, in *Microarrays: Preparation, Microfluidics, Detection Methods and Biological Applications*, Dill, K., R. H. Liu, and P. Grodzinski (Eds.), Springer, New York, pp. 299–314.

Peczalski, A. 2010. Nanowire magnetic compass and position sensor. United States Patent Application 20100024231.

Pelegrí, J., D. Ramírez, and P. P. Freitas. 2003. Spin-valve current sensor for industrial applications. *Sensors and Actuators A: Physical* 105: 132–136.

Perez, J. M., L. Josephson, T. O'Loughlin, D. Hogemann, and R. Weissleder. 2002b. Magnetic relaxation switches capable of sensing molecular interactions. *Nature Biotechnology* 20: 816–820.

Perez, J. M., L. Josephson, and R. Weissleder. 2002a. Magnetic nanosensors for DNA analysis. *European Cells and Materials* 3(Suppl. 2): 181–182.

Perez, J. M., L. Josephson, and R. Weissleder. 2004. Use of magnetic nanoparticles as nanosensors to probe for molecular interactions. *ChemBioChem* 5: 261–264, doi: 10.1002/cbic.200300730

Perez, J. M., F. J. Simeone, Y. Saeki, L. Josephson, and R. Weissleder. 2003. Viral-induced self-assembly of magnetic nanoparticles allows the detection of viral particles in biological media. *Journal of the American Chemical Society* 125: 10192.

Qian, Z., J. M. Daughton, D. Wang, and M. Tondra. 2003. Magnetic design and fabrication of linear spin-valve sensors. *IEEE Transactions on Magnetics* 39(5): 3322–3324.

Reig, C., M.-D. Cubells-Beltrán, and D. R. Muňoz. 2009. Magnetic field sensors based on giant magnetoresistance (GMR) technology: Applications in electrical current sensing. *Sensors* 9: 7919–7942, doi: 10.3390/s91007919

Schellenberger, E. 2010. Bioresponsive nanosensors in medical imaging. *Journal of the Royal Society Interface* 7: S83–S91, doi: 10.1098/rsif.2009.0336.focus

Schellenberger, E. A., F., Reynolds, R. Weissleder, and L. Josephson. 2004. Surface-functionalized nanoparticle library yields probes for apoptotic cells. *ChemBioChem: European Journal of Chemical Biology* 5: 275–279.

Schellenberger, E., F. Rudloff, C. Warmuth, M. Taupitz, B. Hamm, and J. Schnorr. 2008. Protease-specific nanosensors for magnetic resonance imaging. *Bioconjugate Chemistry* 19: 2440–2445.

Srinivasan, B., Y. Li, Y. Jing, Y. H. Xu, X. Yao, C. Xing, and J.-P. Wang. 2009. A detection system based on giant magnetoresistive sensors and high-moment magnetic nanoparticles demonstrates zeptomole sensitivity: Potential for personalized medicine. *Angewandte Chemie—International Edition* 48: 2764–2767.

Tsourkas, A., O. Hofstetter, H. Hofstetter, R. Weissleder, and L. Josephson. 2004. Magnetic relaxation switch immunosensors detect enantiomeric impurities. *Angewandte Chemie* 116: 2449–2453, doi: 10.1002/ange.200352998

Wang, J., W. Xue, N. V. Seetala, X. Nie, E. I. Meletis, and T. Cui. 2006. A micromachined wide-bandwidth magnetic field sensor based on all-PMMA electron tunneling transducer. *IEEE Sensors Journal* 6(1): 97–105.

Xu, L., H. Yu, M. S. Akhras, S.-J. Hana, S. Osterfelda, R. L. White, N. Pourmand, and S. X. Wang. 2008. Giant magnetoresistive biochip for DNA detection and HPV genotyping. *Biosensors and Bioelectronics* 24(1): 99–103, doi: 10.1016/j.bios.2008.03.030

8

Nanobiosensors

8.1 Introduction

The term "biosensor" once referred broadly to any device that responds to chemical species in biological samples or any device meant for biological applications or using biological components. But the definition now is more restricted and sharply focused. A biosensor is a device for the detection of an analyte that combines a biological component with a physicochemical detector. Then simply downscaling the biosensor, a nanobiosensor is a biosensor on the nanoscale size (Khanna 2008).

However, non-scalable parameters ought to be remembered.

Figure 8.1 illustrates the components of a nanobiosensor showing biorecognition, transduction, and interface elements. Molecular recognition is central to biosensing. A nanobiosensor is a simple analytical device having a biological, biologically derived, or biomimetic element utilizing a nanomaterial property or nanoscale phenomenon to convert the recognition event from the analyte into a form suitable for physical or chemical transduction. *What then is the speciality about nanobiosensor?* It brings together the sensitivity and specificity of biological molecules with versatility of physical/chemical transducers at the nano-level. The result is an investigative device whose distinctiveness lies in the fact that it unites the nanotechnological phenomena of physical, chemical, and biological worlds.

What responsibilities are assigned to the various components of a biosensor? The *biological recognition element* or *bioreceptor* of a biosensor ought to be of biological origin such as an enzyme, an antibody (Ab), an antigen (Ag), a cell, a tissue, a deoxyribonucleic acid (DNA). *Do all recognition elements have biological background?* No, not all recognition agents in novel devices are of biological origin, but sometimes they may be synthetically produced elements such as crown ethers, cryptands, calixarenes, and molecularly imprinted polymers. In a wider sense, more correctly, when loosely defined, these artificially produced elements may be accepted and included as components of biosensors as long as biological applications are intended. The recognition element is the component used to fasten the target molecule, and must therefore be highly specific, stable under storage conditions, and immobilized.

The *physicochemical transducer* acts as an interface, measuring the physical change that occurs with the reaction at the bioreceptor, and then transforming that energy into quantifiable electrical output. Finally, the signals from
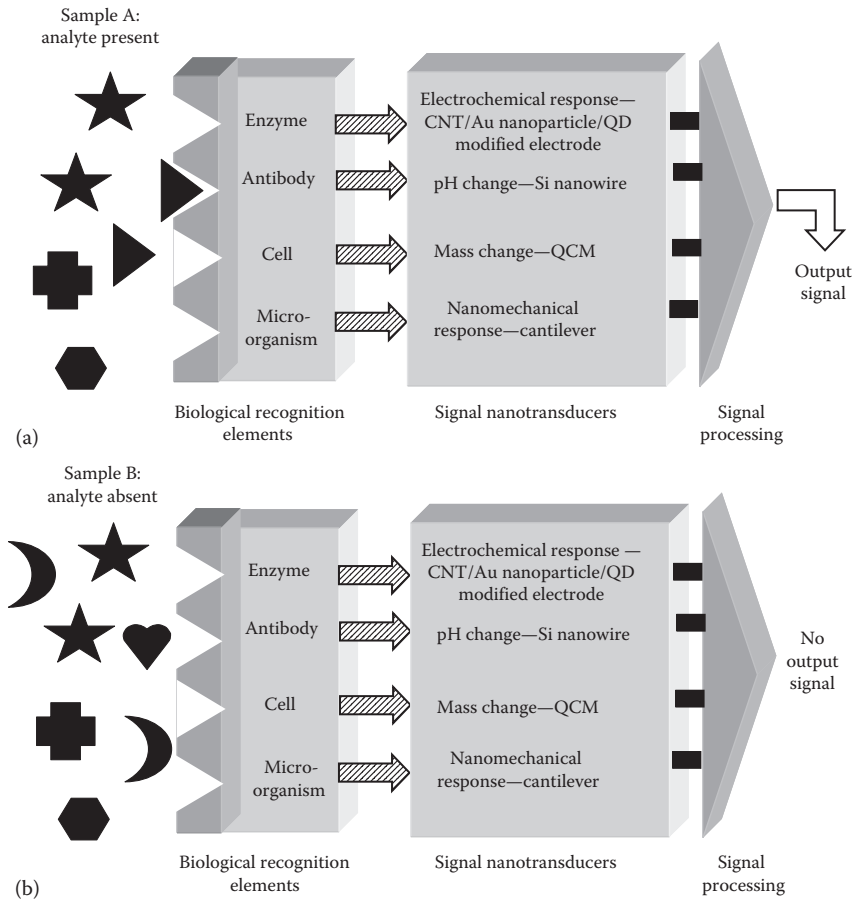
FIGURE 8.1
Nanobiosensor structure: (a) Response of sample A containing analyte and (b) response of sample B having no analyte. In (a), the analyte binds selectively to the bioreceptor yielding a measurable response signal after processing.

the transducer are fed to a *microprocessor* where they are amplified and analyzed; the digitally dissected data is converted to concentration units and transferred to a *display* or/and *data storage device*.

Nanobiosensors are subdivided into five main categories, namely, electrochemical, field-effect, mechanical, optical, and magnetic types. Of these, the first three classes will be exhaustively treated here. Optical nanobiosensors have already been described in Chapter 6 on optical nanosensors and only a cursory view will be presented to append to what has been previously included. Magnetic nanobiosensors have been dwelt upon at length in Chapter 7 on magnetic nanosensors and therefore will not be touched upon here.

8.2 Nanoparticle-Based Electrochemical Biosensors

Because nanomaterials and biomolecules have identical dimensions, the combination of nanometer materials and biomolecules is of considerable importance in the fields of biotechnology and bioanalytical chemistry. Consequent to their large specific surface area and high surface free energy, there has been a growing interest in the study of electron-transfer (ET) between proteins and electrodes modified with nanomaterials (Ju et al. 2002, Wang et al. 2002a and b) since the commencement of the 1990s. Studies have shown that nanomaterial films cannot only offer a welcoming and pleasant platform to assemble protein molecules, but are also able to enhance the ET process between protein molecules and electrodes. Both these qualities have grabbed the attention of researchers.

Which natural element seems to be of paramount importance here? Obviously, it is none other than gold, the ornamental metal in very pure form. Yes, gold has a special role to play among the nanomaterials. Many synthetic procedures are found in the literature for controlling the size, monodispersion, morphology, and surface chemistry of gold nanoparticles (Au NPs). The easy modification of gold surface by thiol-ended molecules makes then suitable for many different biological assemblies.

But gold is a noble metal and bulk gold is nonreactive, famous for retaining its properties and remaining unaffected by the environment. Then how do Au NPs behave differently? Yes, bulk gold is usually catalytically inert in chemical processes. On the other hand, nanometric gold particles have been found to be catalytically active. Gold nanoparticles (GNPs) exhibit remarkable catalytic activity because of their enormous surface area, more availability of reactional sites, and interface-controlled characteristics.

Can metal NPs be used as electrodes? No, it is difficult to directly use metal NPs as electrodes. So, various approaches including self-assembly, grafting reaction electrophoresis etc., were developed to immobilize metal NPs on the surfaces of supporting electrodes so that their useful properties could be utilized. Such electrodes are called *nanoparticle-modified electrodes. How are such electrodes prepared?* Gold NP-modified electrode surfaces are usually prepared in three ways (Yanez-Sedeno and Pingarron 2005): (i) by binding gold NPs with functional groups of self-assembled monolayers (SAMs); (ii) by direct deposition of NPs on the bulk electrode surfaces; and (iii) by incorporation of colloidal gold into the electrode by assimilation gold with other components in the composite electrode matrix.

Modification of surfaces of electrodes used in electrochemistry by decoration with gold NPs bestows on them unusual properties inherent in the nano-modifier Au particles. Several workers have investigated the electrocatalysis of biologically germane small as well as macromolecules at gold-modified electrodes.

What electrochemical techniques have been used for biomolecule sensing? Various electrochemical techniques were used for sensing of biomolecules. Some common techniques are *voltammetry* (an analytical technique for the detection of minute quantities of substances by measuring the currents generated or the substances deposited on electrodes placed in electrolytic solutions, when known voltages are applied), *amperometry* (the measurement of current at a single applied potential), *potentiometry* (measurement of potential under the conditions of zero current flow), *impedometry* (analysis of a sample by impedance measurements), and *conductometry* (determination of the quantity of a material present in a mixture by measurement of its effect on the electrical conductivity of the mixture).

For enhanced current response, it is very important to develop a stable and highly target-specific interface by various surface modifications of conventional electrodes. *Electrocatalysis* is the speeding up of electrode kinetics by a material by minimization of overpotential, the superfluous potential over the equilibrium value that has to be applied to cause an electrodic reaction to take place at a given rate. An *electrocatalyst* is the material, normally employed on the surface or an electrode, which catalyzes a half cell reaction.

Figure 8.2 summarizes basic information on voltammetry techniques. *Linear sweep voltammetry* (LSV) is a voltammetric method where the current at a working electrode is measured while the potential between the working electrode and a reference electrode is swept linearly in time. The *working electrode* is the electrode on which the reaction is occurring. The working electrode is used with an auxiliary electrode and a reference electrode in a three-electrode system. A *reference electrode* is one which has a known, stable electrode potential. *Cyclic voltammetry* (CV) involves application of a triangular potential sweep, allowing one to sweep back through the potential region just covered.

What are electron mediators? How are NPs used as electron wires? In many bioelectrochemical reactions, the electron transfer between the redox-protein and the electrode surface is the key subject to be detected (Lia et al. 2010); a redox protein is one mediating electron transfer in redox reactions. However, the active centers of most oxidoreductases (any of a class of enzymes that catalyze the reversible transfer of electrons from a substrate) are enclosed in considerably thick insulating protein shells. Thus the electron transfer between electrodes and the active centers is blocked by the protein casings. This clogging results in poor analytical performances of electrochemical biosensing without *electron-transfer mediators*. Some compounds, such as ferrocene (the organometallic compound: $Fe(C_5H_5)_2$) derivatives, quinones (either of two isomeric cyclic crystalline compounds $C_6H_4O_2$ having two carbonyl groups CO in an unsaturated six-member carbon ring, biologically important as coenzymes), and poly-2-aminoaniline (PANI) polymer, were able to shuttle electrons giving clearance to current. Common strategies employed to incorporate the redox mediator into the electrochemical system involve either adding the mediator to the solution, or immobilizing
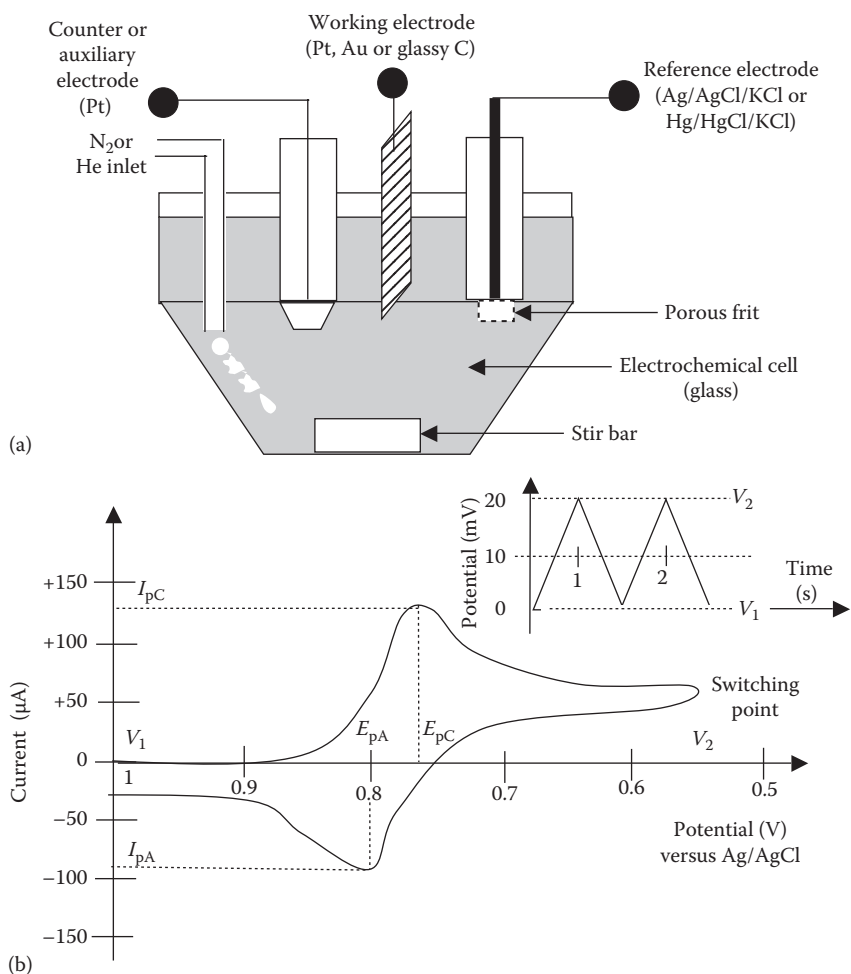
FIGURE 8.2
(a) Electrochemical cell used in voltammetry, consisting of a reference electrode, a working electrode, and a counter electrode along with a nitrogen or helium inlet. (b) Typical cyclic voltammogram of an electrode at a specified sweep rate. I_{pC} and I_{pA} are the peak cathodic and anodic current, respectively, and E_{pC} and E_{pA} the corresponding voltages. (The potential is applied between the reference electrode and the working electrode, and the current is measured between the working electrode and the counter electrode. The working electrode potential ramps linearly versus time, as shown in inset. This ramping is known as the scan or sweep rate.)

it within or on the electrode, producing compacted chemically modified sensors. Immobilization techniques are the dispersion of the mediator in the bulk of a composite electrode or its adsorption, covalent attachment, or polymerization. *How can NPs help here?* Interestingly, the conductivity properties of GNPs can enhance the electron transfer between the active centers

of proteins and electrodes. Thus they act as "electron wires" favoring electron transfer in these applications.

How is the insulating effect of protein shell reduced by gold NPs? Modification of electrode surfaces with gold NPs provides a sympathetic microenvironment similar to that of the redox proteins in native systems, giving the protein molecules more liberty in orientation, thereby reducing the insulating effect of the protein shell for the direct electron ferrying through the conducting passageways or tunnels of gold nanocrystals. It is believed that the nanometric edges of gold particles penetrate slightly into the protein, thereby decreasing the distance between the electrode and the biomolecule redox sites for electron transfer.

8.2.1 Nitric Oxide Electrochemical Sensor

What is the utility of NO sensing? Nitric oxide (NO) plays a decisive role in physiology and pathology. NO has been implicated in the pathogenesis of several diseases; a deficiency of NO owes responsibility in some illnesses (hypertension, hyperglycemia, atherosclerosis, Parkinson's disease, Alzheimer's disease) and conversely, excess NO may participate in others (arthritis, reperfusion injury, cancer). Thus, from biochemical as well as medical perspectives, it is vital to quantify the details of NO production in normal and abnormal tissues, including by direct measurements. Therefore, nanomaterial-based electrochemical sensors selective toward NO have fascinated scientists.

What efforts have been made for NO detection using NP-modified electrodes? Zhang and Oyama (2005) and Goyal et al. (2007) investigated Au-NP arrays directly grown on nanostructured indium tin oxide (ITO) electrodes. They studied the catalytic activity of Au-NP arrays toward the electro-oxidation of NO. SEM (Scanning Electron Microscope) imaging revealed that the size of Au nanospheres could be adjusted by controlling their time of growth, and the as-prepared arrays had good potential use for NO sensing.

Zhang et al. (2008) showed that Pt NPs can be electrodeposited directly on the surface of multi-walled carbon nanotube (MWCNTs) by using a cyclic potential scanning technique to form a novel nano-Pt/MWCNT-modified electrode; the as-prepared Pt-NP showed a potent and powerful electrocatalytic activity for the reduction of NO. Electrocatalysis is the enhancement of electrode kinetics by a material by minimizing the overpotential.

They found that Pt/MWCNT-modified electrode elicited a more sensitive response for the reduction of NO, and it could be developed into a sensor to directly determine the NO concentration in aqueous solution.

How did they examine the response of nano-Pt/MWCNT-modified electrode toward NO? In order to inspect the response character of modified electrode to NO, they studied the detection of NO in phosphate buffer solution with a constant potential voltammetry. A typical amperogram obtained at the modified electrode showed an increase in measured currents with

each addition of 4×10^{-7} mol L^{-1} of NO solution. According to this experiment, a linear relationship between peak currents and NO concentrations was obtained in a range of 4×10^{-7}–1×10^{-4} mol L^{-1} with correlation coefficient $r = 0.997$. The response time (time for the signal increase from 10% to 75%) in the amperometric mode was less than 5 s. The detection limit was estimated to be 1×10^{-7} mol L^{-1}.

Wang et al. (2010) reported a NO electrochemical sensor developed via one-step construction of GNPs–chitosan (CS) nanocomposite sensing film on a glassy carbon electrode (GCE) surface; CS is a linear polysaccharide widely used as an adhesive coating polymer. The NO sensor was prepared by single-step electrochemical deposition in solution containing HAuCl$_4$ and CS on pretreated GCE at an applied potential of –1.00 V for 60 s (GCE/CS–GNPs). Because of special electronic properties and excellent electrocatalytic ability of Au NPs, this method exhibited high sensitivity, wide linearity, and low detection. The anodic peak potential notably shifted negatively compared with that at bare GCE. The high sensitivity and good stability of developed method were coupled to a wide linear range from 3.60×10^{-8} to 4.32×10^{-5} M for the quantitative analysis of NO. The detection limit of 7.20 nM was much lower than the vast mainstream reported methods.

How was the sensor applied to biological samples? Fabricated GCE/CS–GNPs was applied to NO monitoring from rat kidney. No apparent NO current response was observed without kidney or L-Arg in the PBS (Phosphate Buffered Saline) containing trace O$_2$. Arginine, abbreviated as Arg, is an α-amino acid. The L-form is one of the 20 most common natural amino acids.

But a remarkable change of amperometric current could be detected after the addition of L-Arg into PBS containing a whole kidney, suggesting that NO was released from kidney. The concentration of NO liberated from the rat kidney sample was calculated to be ~150 nM. The concentration of NO monitoring from the drug sample was calculated to be ~1.60 μM.

8.2.2 Determination of Dopamine, Uric Acid, and Ascorbic Acid

Why is it necessary to develop biosensors for these analytes? Because dopamine (DA), C$_6$H$_3$(OH)$_2$–CH$_2$–CH$_2$–NH$_2$; uric acid (UA), C$_5$H$_4$N$_4$O$_3$; and ascorbic acid (AA), C$_6$H$_8$O$_6$ are three important biomolecules, which are widely distributed in the body of many mammals, and exhibit vital physiological functions, such as message transfer (communication) in the brain and in defense of the body against disease. DA, one of the major catecholamines, belongs to the family of excitatory chemical neurotransmitters. It is a biogenic (produced by living organisms or biological processes) amine synthesized in the hypothalamus (the part of the brain that lies below the thalamus), in the arcuate (having the form of a bow; curved) nucleus, the caudad (toward the tail or posterior end of the body), and various areas of the central and peripheral nervous systems. Its concentration is of great consequence in the function of central nervous, renal, hormonal, and cardiovascular systems.

Because extreme abnormalities of DA level are warning signs of several diseases such as schizophrenia, Parkinson's disease, the determination of such compounds in real biological samples and identification of changes in neurotransmission that correlate the behavioral states of animals are obvious targets in neurochemical studies, i.e., study of the chemical composition and processes of the nervous system and the effects of chemicals on it. Thus, it is important to develop sensitive, fast, and selective methods for the detection of DA, UA, and AA (Huang et al. 2008).

What are the main obstacles in the correct detection of DA? (i) AA, DA, and UA generally coexist in the extracellular fluid of the central nervous system and serum, and their oxidation potentials are alike at most solid electrodes, which leads to overlapping signals that appear when one strives to detect these three compounds, all together. The major predicament arises in the detection of DA in presence of high concentration of AA. As the AA is oxidized at almost the same potential as DA on the uncovered or bare electrodes, the bare electrode very often is afflicted by the fouling effects due to accretion of the oxidized product on the electrode surface; hence, poor selectivity and sensitivity ensue. (ii) The homogeneous catalytic oxidation of AA by the oxidized DA is another chief interference in the measurement of DA. In view of reasons (i) and (ii), it is very tricky to determine DA directly. In order to resolve these problems, some modified electrodes have been applied to determine DA.

Goyal et al. (2007) applied Au NP arrays for simultaneously detecting DA and serotonin, $C_{10}H_{12}N_2O$. Subsequently, Zhang and Oyama (2007) derivatized the gold NP arrays with 3-mercapto-propionic acid (MPA) whereby a 3D MPA monolayer was produced on the NP array. The 3D MPA monolayer showed stronger electrocatalytic activity toward DA and UA than Au-NP array on ITO.

Huang et al. (2008) deposited a kind of novel nanocomposites, poly(3-methylthiophene) (P3MT)/gold nanoparticles (Au NPs), on the surface of a GCE. 3-Methylthiophene, C_5H_6S, an important conducting polymer, is easily deposited on an electrode surface by an electrodeposition technique. The Au NPs were inserted into a P3MT layer uniformly, and formed a porous 3D structure.

The ability of an Au NPs-P3MT composite–modified electrode for the analysis of DA, UA, and AA at the same time was investigated. This modified electrode showed exceptional electrocatalytic activity toward the oxidation of AA, DA and UA; the overlapping anodic peaks of AA, DA, and UA were completely divided into three well-defined, contrasting voltammetric peaks. A supplementary study showed that there existed a linear relationship between the peak current and the concentration of DA in the range of 1.0×10^{-6}–3.5×10^{-5} mol L^{-1}, and UA in the range of 1.0×10^{-6}–3.2×10^{-5} mol L^{-1}. The detection limits were 2.4×10^{-7} mol L^{-1} for DA and 1.7×10^{-7} mol L^{-1} for UA. This proposed method was applied to the detection of real samples obtaining satisfactory results.

Balal et al. (2009) constructed iron(III)-doped zeolite (one of a family of hydrous aluminum silicate minerals, whose molecules enclose cations of sodium, potassium, calcium, strontium, or barium) NP-modified carbon paste electrode (CPE) for sub-micromolar determination of DA and tryptophan (Trp); Trp is a neutral, genetically coded amino acid, $C_{11}H_{12}N_2O_2$. Differential pulse voltammograms of DA and Trp were compared for zeolite-modified carbon paste electrode (ZCME) and bare CPE. This comparison was made with 30 µM DA and 50 µM Trp in a 0.1 M phosphate buffer at pH = 5. The prepared modified electrode showed voltammetric responses with high sensitivity and stability for DA and Trp.

What are the underlying causes for improved detection capability of DA by NP-modified electrodes? The catalytic mechanism of Au NPs for the oxidation of biomolecules has been studied extensively. Generally, it was considered that the Au NPs layer on the surface of an electrode decreases the overpotential needed, and enable faster ET kinetics, which renders the redox reaction kinetically viable, and the voltammetry appear to be reversible, e.g., with Au NPs on the surface of GCE, the DA oxidation process appeared to be more reversible and the peak shifted amply to be discernible from that of AA, the main interfering factor. The outstanding catalytic ability of the modified electrode was attributed to its special 3D nanostructure, which could accelerate electron transfer; this speeding up produced the observed difference.

8.2.3 Detection of CO

Catalytic oxidation of carbon monoxide has broad technological applications, including fuel-cell technology, air purification in gas products, long-duration space voyage, and CO conversion in automobile exhaust systems. In the experiments of Maye et al. (2000), catalytically active Au NP cores were covered with thiolate (any derivative of a thiol in which a metal atom replaces the hydrogen attached to sulfur; a thiol is a sulfur-containing organic compound having the general formula RSH, where R is another element or radical) monolayers resulting in a networked congregation of core-shell Au NPs on GC electrodes, involving exchange of 1,9 nonanedithiol (NDT), HS-$(CH_2)_9$-SH, with decanethiolate (DT)-capped Au NPs, cross-linking, nucleation and thin film growth; decanethiol has the molecular formula HS–$(CH_2)_9$–CH_3. This assembly had a pronounced catalytic activity toward the oxidation of carbon monoxide with impending sensing applications.

Geng and Lu (2007) studied Au NP on glassy carbon (GC) electrodes for electro-oxidation of carbon monoxide in basic conditions. Electrochemical cyclic voltammetric results illustrated that CO oxidation simultaneously took place in the anodic and cathodic sweeps during one cycle. Moreover, CO electro-oxidation activity was remarkably different in the anode for the different-sized Au NPs. The ultrafine catalyst metal particles (2 and 6 nm Au) were found to be more agile compared to the larger ones (12, 24, and 41 nm Au).

8.2.4 Glucose Detection

Glucose oxidase (GOD) is a structurally rigid glycoprotein (proteins that contain oligosaccharide chains (glycans) covalently attached to polypeptide side chains) having a molecular weight of 152,000–186,000 Da (unit of atomic mass roughly equivalent to the mass of a hydrogen atom, 1.67×10^{-24} g), and consisting of two identical polypeptide chains, each containing a flavin adenine dinucleotide (FAD) redox cenete, which are deeply embedded in the apoenzyme (enzyme that requires a cofactor but does not have one bound); the FAD is a redox cofactor involved in several important reactions in metabolism. Consequently, even if the enzyme approaches the electrode surface, the distance between either of its two FAD/FADH2 centers and the electrode surface stretches beyond the distance across which electrons are transferred at a measurable rate. Naturally, direct electron transfer from the enzyme to the electrode is not observed. In order to shuttle the electrons between the redox centers of GOD and the electrode, ET mediators are often applied, which have been extensively used to fabricate glucose biosensors.

The bioactivity, stability, and quantity of the biological recognition elements immobilized on the electrode are important issues in bioelectrochemistry. Biological activity is a parameter expressing the effects of a molecule on living matter. *How does immobilization of biomolecules on naked surfaces of materials differ from that on NP surfaces?* Generally, the adsorption of biomolecules directly on naked surfaces of bulk materials frequently results in their denaturation and loss of bioactivity. *Denaturation* is a process in which the folding structure of a protein is altered due to exposure to certain chemical or physical factors, e.g., heat, acid, solvents, causing it to become biologically inactive. The GNPs offer excellent candidates for the immobilization platform. The adsorption of biomolecules on the surfaces of GNPs retains their bioactivity and stability because of the biocompatibility and the high surface free energy of GNPs; *biocompatibility* is the property of not producing a toxic, injurious, or immunological response to living matter. As compared with flat gold surfaces, Au NPs, have a much higher surface area, allowing loading of a larger amount of protein and are potentially more sensitive. Thus, a number of laboratories have explored the contribution of GNPs for biomolecular immobilization.

Xiao et al. (2003) reconstituted an apo-flavoenzyme, apo-glucose oxidase, on a 1.4 nm gold nanocrystal functionalized with the cofactor flavin adenine dinucleotide (FAD), a substance that must be associated with an enzyme for it to function. Flavoenzyme is an enzyme that possesses a flavin nucleotide as coenzyme; apo-enzyme is the protein component of an enzyme, to which the coenzyme attaches to form an active enzyme. The electron transfer turnover rate of the reconstituted bioelectrocatalyst was ~5000 s^{-1}. The gold NP acted as an electron relay or "electrical nanoplug" for the alignment of the enzyme on the conductive support and for the electrical cabling of its redox active center. Figure 8.3 shows the procedure of assembling the Au-NP-reconstituted
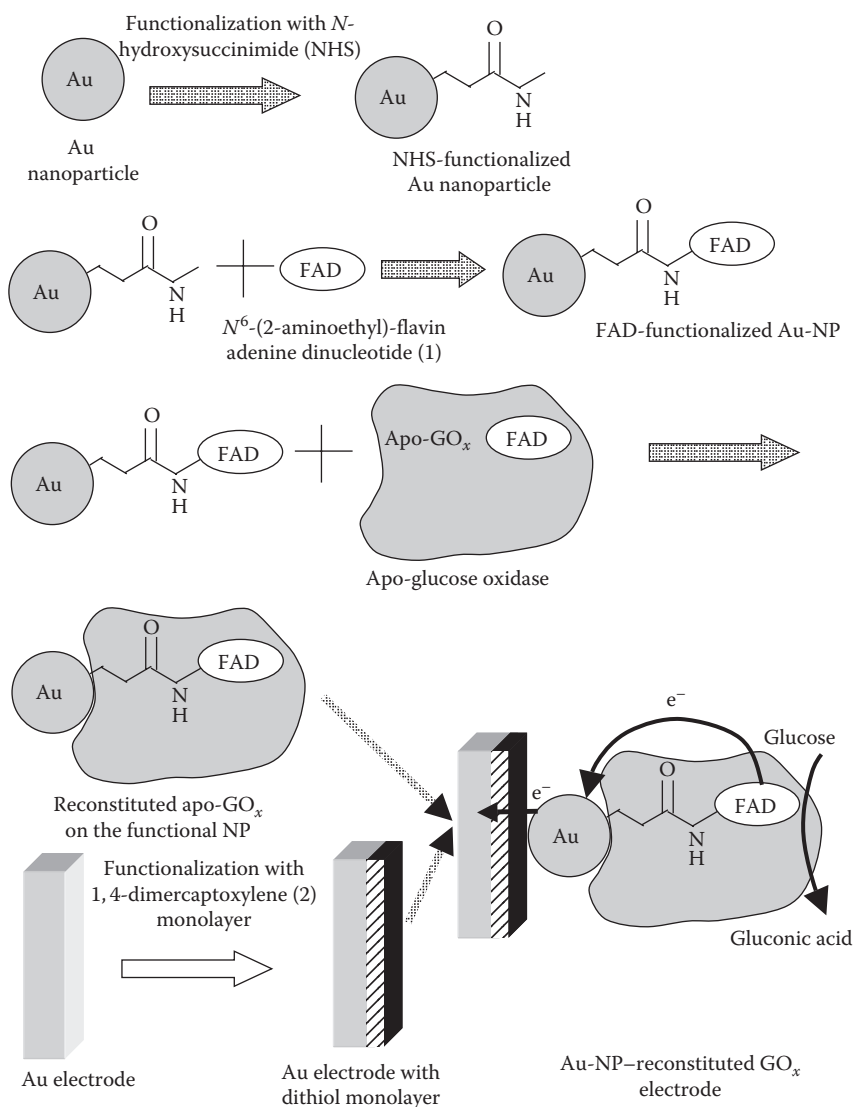
FIGURE 8.3
Steps in the assembly of Au-NP-reconstituted GO_x electrode. (After Xiao, Y., et al., *Science*, 299, 1877, 2003.)

GO_x electrode. For his purpose, Au-NP (1.4 nm) functionalized with a single *N*-hydroxysuccinimide (NHS), $C_4H_5NO_3$, functionality was reacted with N6-(2-aminoethyl)-FAD (1) to yield the FAD-functionalized Au-NP. Apo-GO_x was then reconstituted with the FAD-functionalized Au-NP. The Au-NP-reconstituted GO_x was accumulated on an Au electrode with the FAD-functionalized Au-NP. Thus a strategy was formulated to electrically

contact redox enzymes (GO_x) with their macroscopic environment by the reconstitution of an apoenzyme with an FAD-Au-NP unit.

Andreescu and Luck (2008) developed a sensitive and reagentless electrochemical glucose biosensor based on surface-immobilized periplasmic glucose receptors on GNPs. The sensor was fabricated by immobilization of genetically engineered periplasmic glucose receptors to the GNPs, and showed selective detection of glucose in the micromolar concentration range, with a detection limit of 0.18 µM.

The analytical performances of several enzyme biosensor designs prepared by immobilization of glucose oxidase on different tailored gold NP-modified electrode surfaces were compared by Mena et al. (2005). Glucose oxidase (GO_x) and the redox mediator tetrathiafulvalene (organosulfur compound with the formula ($H_2C_2S_2C)_2$) were co-immobilized in all cases by cross-linking with glutaraldehyde, $CH_2(CH_2CHO)_2$. Gold disk electrodes were modified with short-chain molecules such as cysteamine (Cyst), $HSCH_2CH_2NH_2$, simplest stable aminothiol, and MPA. The 3-mercaptopropionic acid, $C_3H_6O_2S$, and its derivatives are versatile compounds, which are used as chain transfer and cross-linking agents in polymerizations. For Cyst-Au electrodes, the enzyme was immobilized on gold NPs, previously bound on Cyst SAMs using two different configurations. One of them consisted of a Cyst SAM bound to the gold disk via the sulfur atom so that the colloidal gold (Au_{coll}) was linked to the $-NH_2$ moiety (a part or functional group of a molecule). The second one was realized by derivatizing the Cyst SAM with glutaraldehyde to form a thiol terminal group so that gold NPs were bound to the sulfur moiety. Moreover, biosensor configurations based on immobilization of the enzyme onto Cyst or MPA SAMs that were previously bound on electrodeposited gold NPs on GCEs were also evaluated and compared. It was concluded from this comparative investigation that colloidal gold bound to Cyst SAMs gave rise to high rates of enzyme immobilization, and the GOD/Au_{coll}–Cyst–AuE design achieved a much higher stability than for the other configurations. Figure 8.4 presents the scheme for realizing this biosensor design.

The useful lifetime of one single GO_x/Au_{coll}–Cyst–AuE was verified by performing repetitive calibration graphs for glucose after storage of the biosensor in phosphate buffer, pH 7.4, at 4°C. The useful lifetime of GO_x/Au_{coll}–Cyst–AuE was 28 days, extraordinarily longer than that of the other GO_x biosensor designs. The same biosensor yielded slope mean values within the control limits for 28 days. One of the most claimed advantages of the biosensors based on the use of gold NPs is the stability that can be achieved related to the possibility of enzyme immobilization without loss of bioactivity. This performance is amazingly improved when compared with other GO_x biosensor designs.

8.2.5 Gold Nanoparticle DNA Biosensor

Figure 8.5 describes the rudimentary principle of a DNA biosensor. One way of fabricating this on an oxidized silicon surface is by modifying it with
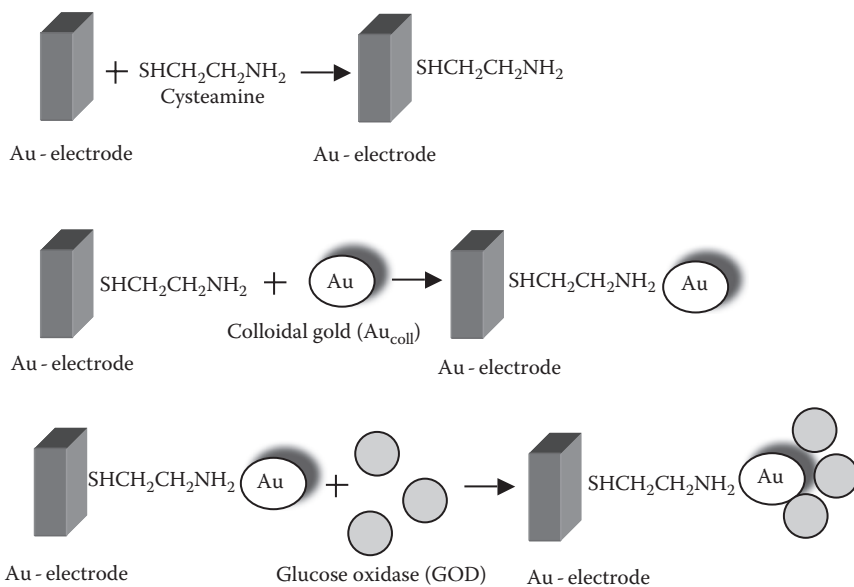
FIGURE 8.4
Preparation of glucose oxidase biosensor designs based on the use of different tailored gold nanoparticle-modified electrode surface: GO_x/Au_{coll}–Cyst–AuE. (After Mena, M. L., et al., *Anal. Biochem.*, 336, 20–27, 2005.)
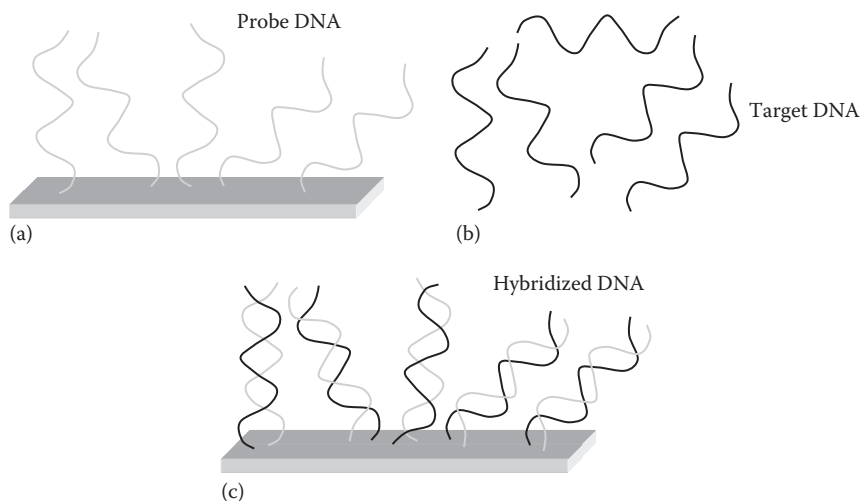
FIGURE 8.5
General DNA biosensor design: Probe DNA in (a) captures target DNA in (b) to form hybridized DNA in (c). The hybridization signal generated is transduced into an electronic signal.

succinimidyl 4-(p-malemidophenyl)-butyrate (SMPB), $C_{18}H_{16}N_2O_6$, a heterobifunctional protein crosslinker reactive toward sulfhydryl and amino groups. Probe oligonucleotides (short polymers of 2–20 nucleotides) are immobilized on the activated surface by spotting NaCl, phosphate buffer (pH = 7) solutions of the appropriate alkylthiol-modified oligonucleotide by manual pipetting. Single-stranded probe DNA is immobilized on the recognition surface. Target capture takes place to generate the recognition signal which is transduced by electrochemical, mechanical, or optical means.

Colloidal Au provides a novel way to construct electrochemical DNA biosensors with high sensitivities for sequence-specific DNA detection (Drummond et al. 2003). Colloidal gold–based electrodes allow the electrode surface area and consequently the amount of immobilized ssDNA to be greatly enhanced. Colloidal gold NPs have been also used to label DNA sequences for the enhancement of signal hybridization, e.g., using a sandwich-based design, the labeled target was captured by probe strands immobilized on a pencil graphite (GP)electrode, and hybridization was subsequently detected by the electrochemical gold-oxidation signal (Ozsoz et al. 2003). The response was greatly enhanced due to the large electrode surface area and the availability of many oxidizable gold atoms in each NP label, the detection limit for PCR (polymerase chain reaction) amplicons being reported to be as low as 0.8 fmol DNA.

In *conductometric DNA detection*, the binding of oligonucleotides functionalized with gold NPs leads to conductivity changes associated with target-probe binding events (Park et al. 2002). The sensitivity of the device was outstandingly improved by exposing the gold NPs to a solution of Ag(I) and hydroquinone when silver deposition facilitated by these NPs bridged the gaps between the particles and led to easily measurable conductivity changes. Figure 8.6 shows the steps involved in this method.

In *electrochemical stripping method* (Wang et al. 2001), after hybridization, the Au NPs were dissolved with HBr/Br$_2$ and related to DNA content. This was achieved by pre-concentration of gold(III) ions through electrochemical reduction and subsequent determination by anodic stripping voltammetry (ASV) (Figure 8.7a). To lower the detection limits, gold tracer "amplification" by silver deposition on the surface was applied (Figure 8.7b).

ASV works by electroplating certain metals in solution on an electrode, thus pre-concentrating the metal. Then the metals on the electrode are sequentially stripped off, generating an easily measurable current. The current (milliamps) is proportional to the amount of metal being stripped off. The potential (voltage in millivolts) at which the metal is stripped off is characteristic for each metal so that the metal can be identified as well as quantified.

Lee et al. (2003) reported an improved electrochemistry-based sequence-specified detection technique by modifying the electrode surface using polyelectrolytes. Polyelectrolytes are polymers whose repeating units bear an electrolyte group. These groups dissociate in aqueous solutions making the polymers charged. Thus they behave both as polymers and electrolytes.
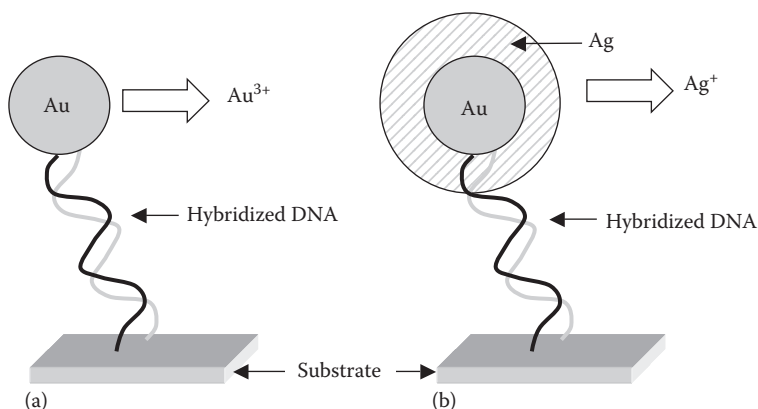
FIGURE 8.6
DNA detection by electrochemical stripping: (a) Au nanoparticles are dissolved with HBr/Br$_2$ with subsequent determination by anodic stripping voltammetry to relate to DNA. (b) Au nanoparticles are amplified by silver deposition followed by stripping analysis.

Standard gold electrode surfaces are vulnerable for significant silver depositions, which creates difficulty in distinguishing whether the electrochemical signal of the silver-dissolution current is from the DNA-bound NP tag or background electrode surface. Lee et al. showed that the background deposition of silver was curtailed by creating a positively charged electrode surface using the electrostatic self-assembly (ESA) of polyelectrolytes. Self-assembly is a process occurring due to spontaneous and uninstructed structural reorganization from a disordered system. ESA occurs when different types of components charge with opposite electrical polarities. The interplay of Coulombic repulsive interactions between like-charged objects and attractive interactions between unlike-charged ones results in the self-assembly of these objects into highly ordered, closed arrays.

Figure 8.8 shows the schematic representation of the electrochemical DNA-hybridization detection. The cleaned gold electrodes were immersed in an ethanolic solution of mercaptoundecanoic acid (MUDA), $HS(CH_2)_{10}CO_2H$ (2 mM) for 48 h at room temperature, again rinsed with water, and dried. The layer-by-layer (LBL) deposition of the polyelectrolytes by alternating the adsorption of the polycations (PAH:poly(allylamine hydrochloride)) and polyanions (PSS:poly(styrenesulfonate)) was performed. LBL is a simple, versatile, and significantly inexpensive approach thin-film fabrication technique for multilayer formation allowing precise thickness control at the nanoscale, in which the films are formed by depositing successively layers of oppositely charged materials with wash steps in between. The outermost negatively charged PSS layer was used to bind the positively charged streptavidin (SA) or avidin (tetrameric protein purified from the bacterium *Streptomyces avidinii*) between PSS and PAH. The immobilization of the DNA probe was accomplished through avidin-biotin binding; biotin is a B vitamin, with
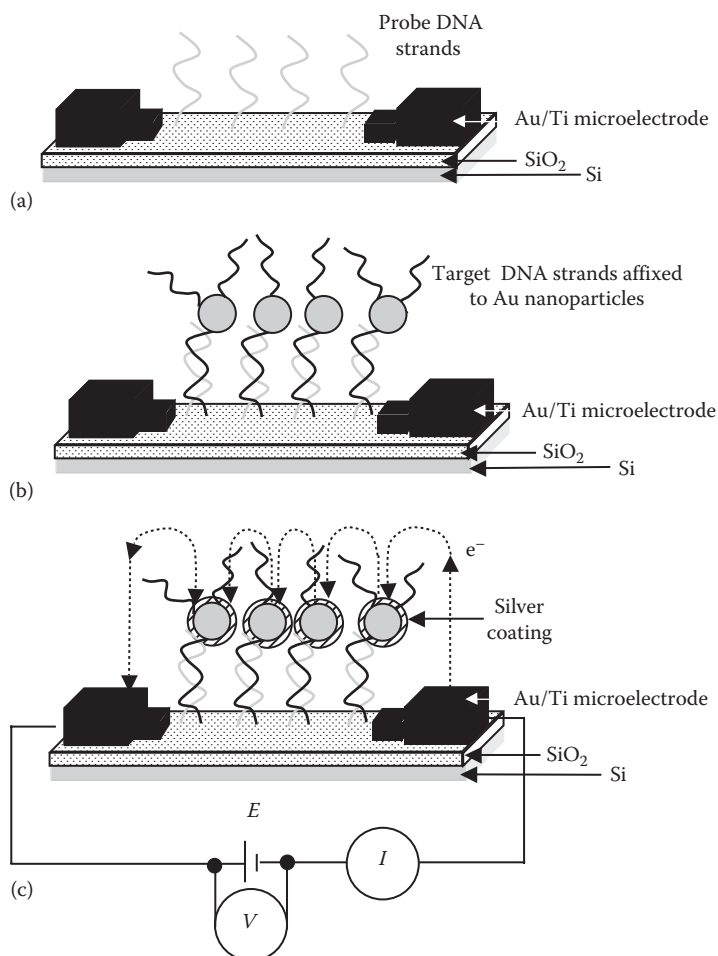
FIGURE 8.7
Conductivity DNA assay: (a) Target DNA strand immobilization on activated silicon dioxide surface. (b) Hybridization of Au nanoparticles tagged oligonucleotides with target DNA. (c) Silver enhancement in $AgNO_3$ and hydroquinone solution, and resistance measurement across electrode gap. (After Wang, J. et al., *Langmuir*, 17, 5739, 2001.)

formula $C_{10}H_{16}N_2O_3S$. The gold electrode was now prepared for hybridization with DNA analytes. The gold NP hybridization indicator was bound to the hybridized target by the interaction between the biotin group of the target and the SA group of the gold NP. Subsequent silver deposition on the gold NP resulted in discernible signal amplification. With the low background signal achieved by electrode-surface modification and selection, sensitive (high signal-to-noise ratio) electrochemical detection of the hybridization event using the silver-enhanced gold NP approach was demonstrated.

8.2.6 Monitoring Allergen–Antibody Reactions

Deriving the advantage of specific recognition between Ab and Ag, through Ab–Ag interaction, that form the basis for detection of infectious disease causing agents, immunoassays are widely applied in clinical practice. Gold NPs have been used in these assays because of their large surface area and biocompatibility. They provide higher loading density and retention of immunoactivity. House dust mites cause heavy atopic diseases such as asthma and dermatitis. Among allergens from *Dermatophagoides farinae*, Der f2 shows the highest positive rate for atopic patients. Huang et al. (2006) assembled recombinant dust mite allergen Der f2 molecules on a gold NP-modified GP carbon
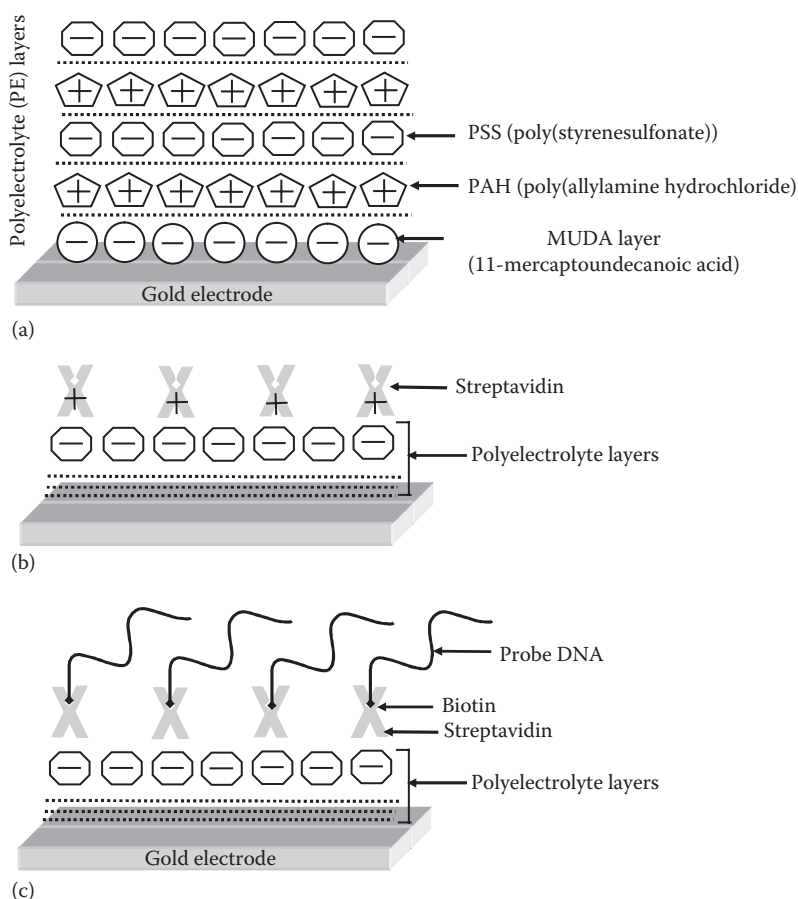
FIGURE 8.8
Electrochemical DNA hybridization detection based on surface modification of gold electrode and silver-enhanced gold nanoparticle labeling: (a) Making polyelectrolyte-modified (PEM) gold electrode. (b) Streptavidin immobilization over the last negatively charged polyelectrolyte layer. (c) Probe DNA layer immobilization over the streptavidin layer.

(d)

(e)

(f)
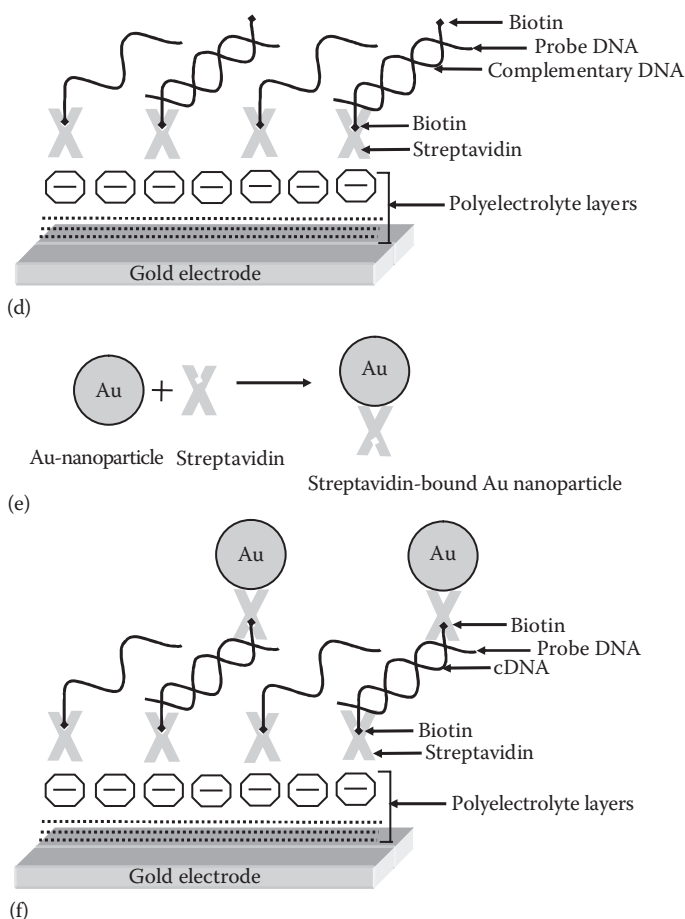
FIGURE 8.8 (Continued)
(d) Target DNA and probe DNA hybridization; the target DNA is biotinylated. (e) Binding of streptavidin with gold nanoparticles. (f) Labeling of hybridized DNA by gold nanoparticles.

electrode and used electrochemical impedance spectroscopy (EIS) to monitor the interaction between the allergen and murine monoclonal antibody (MAb). To make an EIS measurement, a small amplitude signal, usually a voltage between 5 and 50 mV, is applied to a specimen over a range of frequencies of 10^{-3}–10^5 Hz. The EIS instrument records the real (resistance) and imaginary (capacitance) components of the impedance response of the system.

Rise of Ab concentration was accompanied by deceleration of the interfacial electron transference of the redox probe, indicating the bounding of more Ab molecules to the immobilized allergen. Here, immobilization of a larger allergen density with higher immunoactivity retention was achieved due to the active sites of Au colloidal monolayer. In EIS, the gold NPs served as electrical conduction paths.

Au

Au — Silver

Biotin
Probe DNA
cDNA

Biotin
Streptavidin

Polyelectrolyte layers

Gold electrode

(g)

$Ag \rightarrow Ag^+ + e^-$

Au

Au

A

Biotin
Probe DNA
cDNA

Biotin
Streptavidin

Polyelectrolyte layers

Gold electrode

(h)
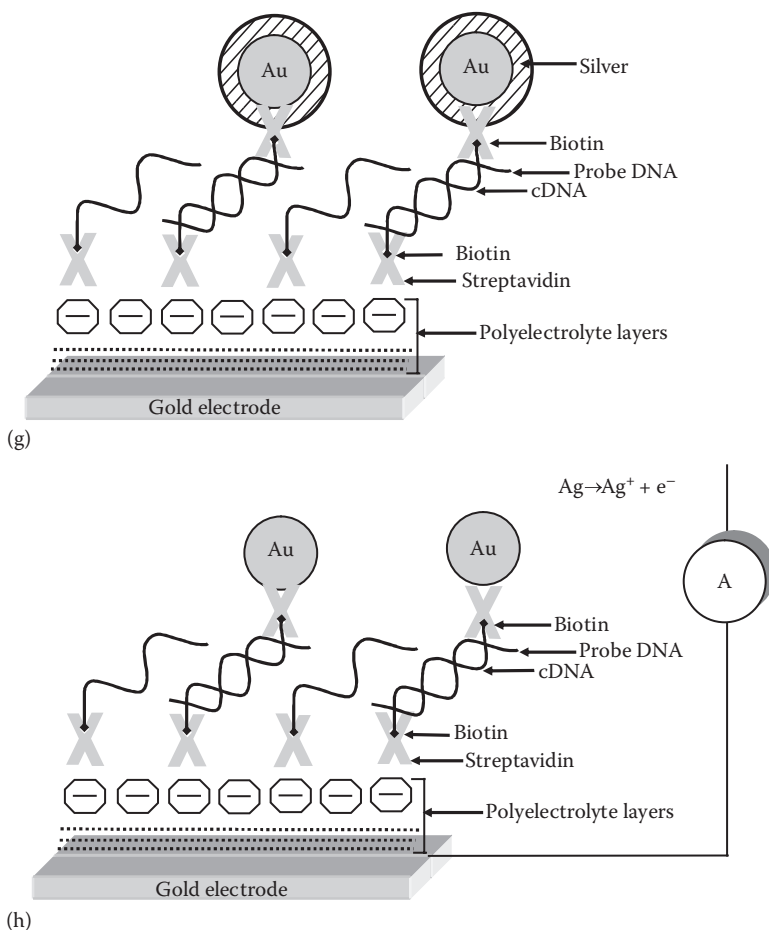
FIGURE 8.8 (Continued)
(g) Staining of gold nanoparticles with silver. (h) Measuring oxidative silver-dissolution current. (After Lee, T. M.-H. et al., *Langmuir*, 19, 4338, 2003.)

8.2.7 Hepatitis B Immunosensor

The binding reaction of Ags at Ab-immobilized surfaces is often insufficient to produce a large signal change, e.g., for impedance measurements. A novel strategy used to overcome this shortcoming is the immobilization of Ab molecules via self-assembly on a Pt electrode modified with colloidal gold and polyvinyl butyral, $H_2(C_8H_{14}O_2)_n$… (PVB: a tough, clear, adhesive) as matrices, which combines the large specific surface area and high surface free energy of Au NPs with the entrapped effect of PVB (Tang et al. 2004). The novel immunosensors exhibited excellent electrochemical characteristics toward Hepatitis B surface antigen, (HBsAg),

the most frequently used to screen for the presence of this infection, which appears in the serum from 4 to 12 weeks following infection. The method is applied generally to amplified assays of other biomolecules and other transduction modes, such as potentiometric or amperometric immunosensors.

8.2.8 Carcinoembryonic Antigen Detection

Most of the works involving immunosensors are related to the diagnosis and/or monitoring of human diseases. The carcinoembryonic Ag is a glycoprotein used as tumor marker and has been frequently investigated in immunoreactions. It is present in fetal gastrointestinal tissue, but generally absent from adult cells with the exception of some carcinomas, malignancies that originate in the epithelial tissues.

A label-free amperometric immunosensor was fabricated by Ou et al. (2007) on a mercaptopropanesulfonic acid ($C_3H_8O_3S_2$)-modified gold electrode surface based on LbL (LBL) assembly of Au NPs, multi-walled carbon nanotubes-thionine (MWCNT-Thi) and CS (Chit), (AuNPs/ MWCNT-Thi/ Chit)$_n$, and posterior anti-carcinoembryonic antigen (CEA) immobilization via covalent bond; thionine has the formula Cl_2HSNaS, CEA stands for Carcinoembryonic Ag. The detection was based on the variation of current responses before and after the immunoreaction. When the immobilized antibodies have bound with Ags, the Ag-Ab complex formed on the surface inhibited the ET. Since the formed film by LBL presents an electroactive (exhibiting a change in size or shape when stimulated by an electric field) profile, the modified electrode amperometric signal was modulated by subsequent addition of target molecules. It was verified by a decrease of the amperometric signal as the concentration of Ag on surface increased.

8.2.9 *Escherichia coli* Detection in Milk Samples

Lin et al. (2008) described the detection of *Escherichia coli* in milk samples. *E. coli* is a bacterium that is commonly found in the gut of humans and other warm-blooded animals, one of the most frequent causes of many common bacterial infections, including cholecystitis, bacteremia, cholangitis, urinary tract infection (UTI), and traveller's diarrhea, and other clinical infections such as neonatal meningitis and pneumonia.

In this work, a disposable amperometric immunosensor was developed by using a screen-printed carbon electrode modified with ferrocenedicarboxylic acid ($C_{11}H_{10}FeO_2$) and AuNPs which lead to increase in the amount of anti-*E. coli* immobilized on electrode surface achieving high sensitivity. The anti-*E. coli* was immobilized via cross-linking by using glutaraldehyde and after adsorption of Ag, a second horseradish peroxidase (HRP)-labeled Ab (Ab conjugated to HRP) was immobilized.

8.3 CNT-Based Electrochemical Biosensors

In what ways do one-dimensional nanostructures aid in bioelectronic sensing? Importance of one-dimensional nanostructures like CNTs for bioelectronic detection stems from their electronic transport properties besides their high surface-to-volume ratio and the capacity to bunch a tremendous density of sensing elements on the footprint of a small array device. Carbon nanotubes work as efficient substrates or modifiers for promoting ET reactions. They exhibit excellent electron transfer supporting ability when used as electrode modifier in electrochemical reaction electrodes. This is due to their high electronic conductivity for the electron transfer reactions. Over and above, they have better electrochemical and chemical stabilities in both aqueous and nonaqueous solutions. They also offer easy protein immobilization with retention of activity for biosensors.

Where actually are the electroactive sites located in CNTs? Holloway et al. (2008) studied the voltammetry of two standard redox processes of $Fe(CN)_6^{4-}$ and $Ru(NH_3)_6^{3+}$ complex ions using three types of MWCNTs having oxygenated edge-plane, edge-plane, and almost no edge-plane like defects. It was found that the electrochemistry of CNTs was dominated by *the ends of the tubes*. The rate of electron transfer was faster in the case of *edge-plane defect sites*. This indicated that the electroactive sites on MWCNTs were located at the tube ends. It was also observed that aligned SWCNT-modified electrodes showed better electrochemical properties authenticating further that the electrochemistry of CNTs was dominated by the ends of the tubes. The length of the aligned CNTs also had a significant effect on the electron transfer rate. The electron transfer rate constant was inversely proportional to the mean length of the CNTs. The electron transfer rate at the CNT-modified electrode was further increased by dialyzing the CNTs after purification and shortening treatment. The adsorbed acid moieties during purification and acid-treatment processes decreased the electrocatalytic activity of CNTs in electroanalysis.

How are CNT-modified electrodes made? CNT-biocomposite electrodes consisting of a mixture of CNTs, binders, and biomolecules, are made by mixing carbon powder with different binders such as mineral oil or bromoform ($CHBr_3$). *What are the advantages of such electrodes?* CNT-modified electrodes offer advantages such as high aspect ratios ~100–1000, nano sizes, specific catalytic activity together with sensitivity, selectivity, stability, and reliability.

How do CNTs help in communication between active sites of enzymes and the electrode? Establishing a satisfactory communication channel between the active site of the enzyme and the electrode surface is a formidable problem in amperometric enzyme electrodes. The enzyme cannot be oxidized or reduced at the electrode at any potential because it is embedded deep inside the shell. The ability of CNTs to reduce the distance between the redox site of the protein and the electrode helps in accelerating the overall rate of the reaction, the rate of electron transfer varying inversely with the exponential

separation between the redox center and the electrode. The small diameter and long length of CNTs allow them to be plugged into proteins with better electro-activity compared to other carbon-based electrodes. Patolsky et al. (2004) have reported the direct electrochemistry of enzymes at CNT-modified electrodes.

8.3.1 Oxidation of Dopamine

Sincere effort has been made in the development of a highly sensitive and selective method for the detection of DA, which is one of the important catecholamine ("fight or flight" hormone released by the adrenal glands in response to stress) neurotransmitters in the mammalian central nervous system. Abnormal concentrations of DA in body fluids influence the function of central nervous system. *Why conventional electrodes are unable to detect DA?* Conventional electrodes are not suitable for the determination of DA due to the interference from AA and UA, which coexist in a real sample at 100 times higher concentration than DA. These compounds can be easily oxidized at the similar potential of DA and thus always interfere with DA detection.

How do CNT-modified electrodes resolve this problem? CNT-modified electrodes are widely used for DA sensing. A CNT electrode was constructed by Britto et al. (1996) by mixing CNTs with bromoform and packing the paste inside a glass tube. Oxidation of DA was studied by CV with the constructed CNT-modified electrode. CV is a type of potentiodynamic electrochemical measurement having a current peak on the forward scan and a second, inverted peak on the reverse scan representing the opposite reaction (oxidation or reduction) to that observed on the forward scan. CV takes the experiment a step further than LSV, which ends when it reaches a set potential. When CV reaches a set potential, the potential ramp of the working electrode is inverted. This inversion can happen multiple times during a single experiment.

A peak potential separation of 30 mV was obtained reversibly at the CNT electrode which was far superior to that at the CPE. Due to the catalytically active surface, background currents were larger in the former case. Voltammetric behavior of DA was unaffected by the treatment of CNT electrode with homogenized brain tissue suggesting that DA was oxidized mainly inside the tubes where the electrolytic product was stabilized. The enhancement in performance by the CNTs was attributed to the abundance of active sites on the tube surfaces and ends, their specific electronic structure, and the presence of edge-plane like defects on the surfaces of CNTs.

Wang et al. (2006) reported the fabrication of a P3MT-modified GCE coated with Nafion/SWCNTs for highly selective and sensitive determination of DA. P3MT, C_5H_6S, is a widely used conducting polymer. The modified electrode enhanced the voltammetric signal of DA and effectively suppressed the interferences of AA and UA as well. A lower detection limit of 5.0 nM was achieved for DA. It was also successfully applied for the determination of DA in healthy human blood serum.

Goyal et al. (2008) used a fullerene-C60 coated gold electrode for the determination of DA in presence of high concentration of AA at physiological pH; fullerene is any molecule composed entirely of carbon, in the form of a hollow sphere, ellipsoid, or tube. Since the discovery of fullerenes in 1985, buckminsterfullerene (C60) has ensnared a large number of scientists due to its remarkable electrochemical properties. The modified electrode not only exhibited strong catalytic activity toward the oxidation of DA and AA, but also separated their voltammetric responses in the presence of each other, and thus DA was detected selectively in the presence of excess of AA. Linear calibration curves for DA were obtained using square-wave voltammetry over the concentration range 1 nM–5.0 mM in 0.1 M phosphate buffer solution at pH 7.2 with a correlation coefficient of 0.9931 and the detection limit (3σ: an upper or lower control limit 3 standard deviations from the mean) was estimated to be 0.26×10^{-9} M.

The interference studies showed that the presence of physiologically common interferents (i.e., UA, citric acid, tartaric acid, glucose, and sodium chloride) negligibly affected the response of DA. On the other hand, the bare electrode was incapable in separating the responses of DA and AA. AA caused major interference in the voltammetric determination of DA. A somewhat broad oxidation peak was obtained and the oxidation peak potentials of the analytes were indistinguishable.

8.3.2 Direct Electrochemistry or Electrocatalysis of Catalase

Wang et al. (2002, 2004) and Li et al. (2003) prepared a CNT-modified electrode by casting an aliquot (a portion of a total amount of a solution) of a CNT suspension on the substrate surfaces. It was found that the as-prepared electrode facilitated the ET reactions of cytochrome *c* (cyt *c*), catalase (*Ct*), and DNA. cyt *c* is defined as electron transfer protein having one or several heme *c* groups, bound to the protein. Heme is a prosthetic group (a tightly bound, specific non-polypeptide unit required for the biological function of some proteins) that consists of an iron atom contained in the center of a large heterocyclic organic ring called a *porphyrin*. *Ct* is an enzyme found in most plant and animal cells that functions as an oxidative catalyst.

Cyclic voltammograms of *Ct* were recorded at a bare gold electrode and at a CNT-modified electrode. Within a typical potential window, no voltammetric response was observed for *Ct* at the bare gold electrode whereas a quasi-reversible redox process of *Ct* was obtained at the modified electrode, corresponding to the Fe(III)/Fe(II) redox center of the heme group of the *Ct* adsorbate. It was noted that the electron transfer rate of the *Ct* redox reaction was much faster at the SWCNT-modified electrode with respect to other electrodes based on carbonaceous materials (substances rich in carbon) like carbon soot and GP. Moreover, the catalytic activity of *Ct* was retained upon adsorption on the modified electrode.

8.3.3 CNT-Based Electrochemical DNA Biosensor

DNA analysis is helpful to understand many diseases on a molecular level and promises new perspectives for medical diagnosis in future. DNA bearing the genetic information is a double helix molecule and the double helix is held together by two sets of forces: hydrogen bonding between complementary base pairs and base-stacking interactions.

What is the principle of electrochemical hybridization biosensors? In these biosensors, a single-stranded DNA probe is immobilized on the transducer surface. This converts the formation of a double-stranded DNA into the required electrical signal.

(i) The label-free approach works on the intrinsic redox-active properties (e.g., direct oxidation) of the DNA bases: guianine or adenine. (ii) A redox-active-labeled system utilizes a redox active label and there is a change in affinity of the redox molecule toward the probing single-stranded-DNA-modified interface before and after exposure to target DNA. In both the approaches, it is desirable that the current signal originating from the hybridization between the probe and the target be properly amplified.

How is this amplification done? In the label-free method, the guanine oxidation is amplified by using MWCNTs. Performing direct electrochemistry of guanine at a CNT-modified electrode yields a comparatively higher guanine oxidation current than an unmodified GCE using CV. For the detection of DNA sequences related to breast cancer *BRCA 1* gene using CV analysis, the guanine oxidation current observed for MWCNT-modified electrode was 11 times that with an unmodified GCE (Wang et al. 2003a and b). Kerman et al. (2004) employed sidewall and end-functionalized CNTs providing a larger surface area for DNA immobilization and higher electron transfer efficiency to achieve further increase in guanine oxidation current. The detection limit for DNA was also lowered by this method.

8.3.4 Glucose Biosensor

What is the concept of molecular wiring? How are CNTs used for this purpose? Electron transfer in biological systems is one of the leading areas in the biochemical and biophysical sciences. In the past few years, there has been considerable interest in the direct electron transfer between redox proteins and electrode surfaces. Various immobilization strategies have been adopted to fabricate enzyme electrodes for biosensor applications. Evidently, the best strategy for successful enzyme biosensor fabrication is to devise a configuration by which electrons can directly transfer from the redox center of the enzyme to the underlying electrode. This has been accomplished in recent years using the idea of *molecular wiring*.

The resemblance in length scales between nanotubes and redox enzymes suggests the presence of interactions that may be encouraging for biosensor electrode applications. The strategy of physical adsorption or covalent

immobilization of large biomolecules on the surface of immobilized carbon nanotubes may epitomize an exhilarating pathway through which direct electrical communication between electrodes and the active site of redox-active enzymes is achieved. The nanotubes are dispersed on the surface of a support electrode to form a haphazardly sprinkled array of high surface area, or else incorporated as a dispersion within a matrix to form a thin film.

How is molecular wiring done? Patolsky et al. (2004) reported the structural alignment of the enzyme glucose oxidase, GO_x, on electrodes by using SWCNTs as electrical connectors between the enzyme redox centers and the electrode. To achieve this, SWCNTs were purified and shortened. The shortened SWCNTs were further purified by dialysis and filtering. The length-fractionalized SWCNTs were coupled to the surface of an Au electrode on which a 2-thioethanol (C_2H_6OS)/cystamine (organic disulfide: $C_4H_{12}N_2S_2$) mixed monolayer (3:1 ratio) was assembled. As depicted in Figure 8.9, this coupling was done in the presence of the reagent 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC or EDAC), a zero-length cross-linking agent used to couple carboxyl groups to primary amines. A densely packed, needle-like pattern of standing SWCNTs was obtained after 5h of coupling, leading to a preferred standing conformation (the spatial arrangement of atoms and chemical bonds in a molecule) of the SWCNTs on to the surface. The amino derivative of the FAD cofactor (1), was then coupled to the carboxyl groups, –COOH, at the free edges of the standing SWCNTs. After this step, apo-glucose oxidase (apo–GO_x) was reconstituted on the FAD units linked to the extremities of the standing SWCNTs. The study demonstrated the aligned reconstitution of a redox flavoenzyme (glucose oxidase) on the edge of carbon nanotubes that were linked to an electrode surface. The SWCNT acted as a nanoconnector that electrically contacted the active site of the enzyme and the electrode. The electrons were transported along distances greater than 150 nm, and the rate of electron transport was controlled by the length of the SWCNTs.

Lyons and Keeley (2008) studied carbon nanotube–based modified electrode biosensors. Among electrochemical biosensors based on enzyme attachment, unquestionably, the glucose oxidase (GO_x) assemblies are the most deeply studied systems. This enzyme has attractive characteristics such as its well-known behavior, great stability, and sturdiness. Comparison of the voltammograms obtained using a bare, SWCNT-modified, and SWCNT/GO_x and SWCNT/GO_x/Nafion-modified GC working electrodes in a 50 mM phosphate buffer (pH = 7) revealed that a bare GC electrode exhibited a virtually flat and uninspired, featureless voltammetric response, whereas a pair of well-defined redox peaks was observed at the both the SWCNT/GO_x and SWCNT/GO_x/Nafion-modified electrode. Nafion ($C_7HF_{13}O_5S \cdot C_2F_4$) is a sulfonated tetrafluoroethylene-based fluoropolymer-copolymer.

The observed peaks were characteristic of those representing the redox behavior of an adsorbed species, that of the FAD. The summits were
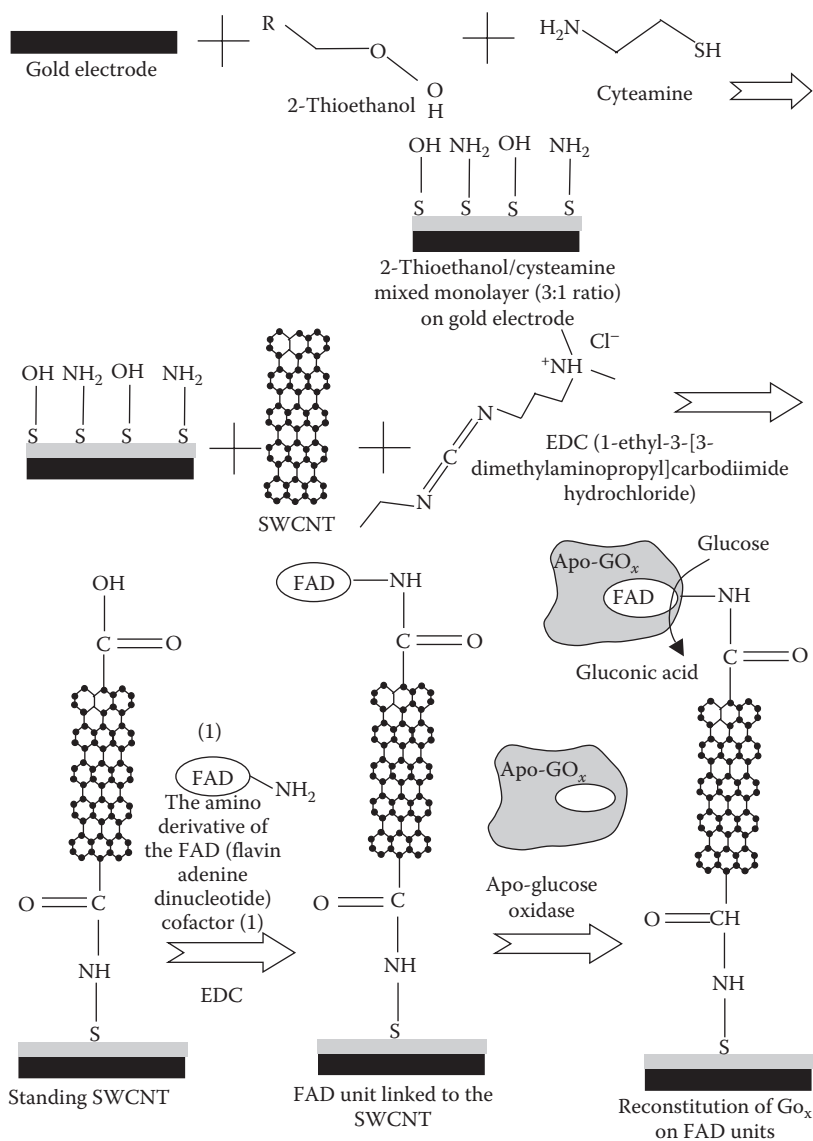
FIGURE 8.9
SWCNT electrically contacted glucose oxidase electrode. (After Patolsky, F. et al., *Angew. Chem. Int. Ed.*, 43, 2113, 2004.)

compellingly symmetrical and exhibited the characteristic bell shape expected for adsorbed redox species. Further, it was noted that that the addition of Nafion clearly enhanced the dispersion of the nanotube collections on the support electrode surface, which may then boost the effectiveness of the nanotube wiring to the entrenched flavin active site of the immobilized

glucose oxidase, and hence result in more successful charge transfer kinetics and potential directed turnover of the flavin group. Nafion-overcoated nanotube/GO_x composite-modified electrode was unmistakably superior to GC as an enzyme immobilization platform.

What explanations have been proposed for improved response of modified electrodes? These are enumerated as follows:

1. Wang et al. (2003) have reported that glucose oxidase adsorbs preferentially to edge-plane sites on nanotubes. It has been established that such sites contain a significant amount of oxygenated functionalities. These groups are created via the rupturing of carbon–carbon bonds at the nanotube ends and at defect sites, which may occur on the side walls. They render hydrophilicity (water loving: the tendency of a molecule to be solvated by water) and ionic character to the nanotubes, and it is believed that they are responsible for the "nesting" of the protein on the nanotube film. The nanotubes and enzyme molecules are of similar dimensions, which makes possible the adsorption of glucose oxidase without major loss of its biocatalytic shape, form or function.

2. Lyons and Keeley (2008) suggested that nanotubes pierce the glycoprotein shell and gain access to the prosthetic group such that the electron tunneling distance is curtailed. Such access is not generally likely with conventional "smooth" electrodes, and notable unfolding of the protein shell can occur, resulting in the loss of biochemical activity. Further, they proposed that the GO_x molecules are adsorbed in a plausibly pristine configuration where the tunneling distance for electron transfer between the flavin sites and the nanotube strand is not too large; hence, making the transition probability for electron transfer favorable. It is important to note that the carbon nanotube and the enzyme molecule share a similar length scale and so the enzyme is able to adsorb on the nanotube sidewall without losing its biologically active shape, form, and function.

3. Guiseppi-Elie et al. (2002) have introduced the striking analogy of piercing a balloon with a long sharp needle such that the balloon does not burst. Instead by a gentle twisting action, the needle is made to enter the balloon uncatastrophically. Similarly, it has been proposed that some number of nanoscale "dendrites" of CNT venture outward from the surface of a strand and act like bundled ultramicroelectrodes that are able to perforate the glycoprotein shell of glucose oxidase and gain access to the flavin prosthetic group such that the electron tunneling distance is minimized and consequently, electron transfer probability is optimized. This degree of nanoscale electronic wiring and intimate access is not generally afforded with traditional smooth electrodes.

8.3.5 Cholesterol Biosensor

Normal human blood serum contains less than 200 mg dL^{-1} cholesterol. A waxy, fat-like substance that the body needs to function normally, it is naturally present in cell walls or membranes. Li et al. (2005) reported a carbon nanotube-modified biosensor for monitoring total cholesterol in blood; the schematic of the sensor is shown in Figure 8.10. They used carbon nanotubes to modify CPE, which promotes electron transfer and enhances the response current. This sensor consists of a carbon working electrode and a reference electrode screen-printed on a polycarbonate (a versatile, tough plastic: polymer containing carbonate groups (–O–(C=O)–O–), produced by the reaction of bisphenol A and phosgene) substrate. *What are the ingredients of the mixture applied on the electrodes?* Cholesterol esterase (an enzyme release cholesterol and a fatty acid anion), cholesterol oxidase (an alcohol dehydrogenase/oxidase flavoprotein catalyzing the dehydrogenation of C(3)-OH of cholesterol), peroxidase (HRP is hydrogen-peroxide-reducing enzyme, occurring in animal and plant tissues, that catalyzes the dehydrogenation (oxidation) of various substances in the presence of hydrogen peroxide), and potassium ferrocyanide, $K_4[Fe(CN)_6]\cdot 3H_2O$, were immobilized on the screen-printed carbon electrodes. MWCNTs were added to prompt electron transfer. The role of redox mediator potassium ferrocyanide was to make the reaction detectable at low potential
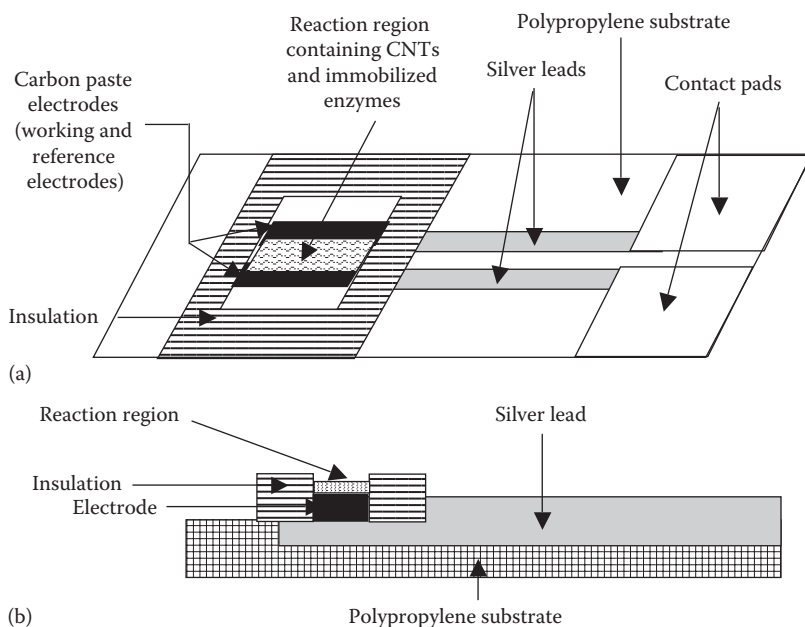
FIGURE 8.10
Cholesterol biosensor consisting of two screen-printed carbon paste electrodes: (a) the sensor strip and (b) its cross section. (After Li, G. et al., *Biosens. Bioelectron.*, 20, 2140, 2005.)

to minimize the effect of interferences. The function of the silver leads was to improve electric conductivity of the electrodes. The reaction area, where the carbon paste film was modified by carbon nanotubes and immobilized enzymes, was defined by the insulating film coated on the carbon paste film. There were 20 sensor bases on each plate. Once the sensor array was complete, sensor strips were incised out of the array to be measured. The size of the reaction area was $2\,mm \times 2\,mm$ while the sensor strip was $35\,mm \times 10\,mm$.

By how much do the responses of modified and unmodified electrodes differ? The responses to cholesterol were determined over the range of $100-400\,mg\,dL^{-1}$ using both the carbon nanotube modified and unmodified sensors. An almost linear relationship between the cholesterol concentration and the response current of the carbon nanotube–modified electrodes was observed, while a curved response was observed for unmodified electrodes. The average sensitivities were 0.0059 and $0.0032\,\mu A\,mg^{-1}\,dL^{-1}$, respectively. It is seen that the carbon nanotubes promote the electron transfer and almost double the sensitivity. AA and UA at normal levels expected in blood had no effect on the response current. This indicates that these substances did not interfere during the blood total cholesterol measurements.

8.3.6 H_2O_2 Biosensor

Chen et al. (2007) have fabricated a H_2O_2 biosensor based on the immobilization of hemoglobin (Hb) on multiwall carbon nanotubes and gold colloidal NPs. Liu and Hu (2008) developed a similar analytical device by entrapping Hb in a composite electrodeposited CS-multiwall carbon nanotubes film by assembling gold NPs and hemoglobin step by step. Hb is a kind of redox respiratory protein in red cells. It is the iron-containing oxygen-transport metalloprotein in the red blood cells of all vertebrates except fishes. CS is a form of natural polysaccharide. GNPs provide a mild microenvironment similar to that of redox proteins in native systems and give the protein molecules more freedom in orientation.

How is the hemoglobin-modified electrode prepared and what are its characteristics? The CNTs were dispersed in CS solution to obtain a homogeneous suspension. Then the suspension was electrodeposited on to the surface of gold electrode, and the modified electrode was transferred to GNPs solution for assembling. Finally, hemoglobin was adsorbed onto the GNPs. The composite electrode showed excellent electrocatalysis to hydrogen peroxide and oxygen. The peak current was linearly proportional to H_2O_2 concentration in the range from 1×10^{-6} to $4.7 \times 10^{-4}\,mol\,L^{-1}$ with a detection limit of $5.0 \times 10^{-7}\,mol\,L^{-1}$. This biosensor exhibited high stability, good reproducibility, and selectivity. This is due to the synergistic action of both carbon nanotubes and Au NPs, since both of them have the ability to facilitate or promote the electron transfer between the proteins and the electrode surface. The incorporation of CNTs and GNPs in the film increased its electrocatalytic ability toward hydrogen peroxide.

8.4 Functionalization of CNTs for Biosensor Fabrication

There are two principal methods of CNT functionalization: (i) non-covalent and (ii) covalent. Non-covalent methods include encapsulation, physical and chemical adsorption, and hydrophobic interactions. Covalent immobilization of biomolecules on CNTs is implemented by oxidation of CNTs by sonicating (shearing open by high-frequency sound agitation, usually ultrasound) or refluxing (to boil a liquid in a vessel attached to a condenser so that the vapors continuously condense for reboiling) in concentrated acid solution which results in the formation of carboxylic acid –COOH, groups at the ends and sidewalls if the nanotubes. Carboxylic acid groups on the surfaces of CNTs react with amino functional groups (RNH$_2$ or R$_2$NH, or R$_3$N where R is a carbon containing substituent) of biological receptors through carbodiimide (or methanediimine is a functional group of the formula RN=C=NR formed by dehydration of urea) coupling, e.g., the attachment of bovine serum albumin (BSA) protein (a serum albumin protein used as a carrier protein and as a stabilizing agent in enzymatic reactions) and glucose oxidase (GOD) enzyme on the sidewalls of CNTs via amide (an organic compound that contains the functional group consisting of an acyl group (R–C=O) linked to a nitrogen atom (N)) linkages.

   Most research studies on CNTs have employed the *droplet coating technique* for preparation of CNT electrodes. Here, the CNTs are randomly aligned with respect to each other and are pointing in different directions, resulting in a tangle of nanotubes. An unknown spatial relationship exists between the biomolecules and the CNTs. But for biomolecules and CNTs to operate in harmony, attachment of biomolecules to CNTs needs cautious attention. In a CNT, there are two types of sites for immobilization of biomolecules. These attachment sites are on the tips and sidewalls of the nanotubes. Tip conjugation is favored in densely packed arrays because sidewalls provide limited accessability. But a large surface area is available for conjugation on the sidewalls of CNTs because of their high aspect ratio. So, sidewalls cannot be ignored. In many applications such as a CNT forming the channel of a MOSFET (metal-oxide-semiconductor field-effect transistor), sidewall conjugation is the solo possibility.

8.5 QD (Quantum Dot)-Based Electrochemical Biosensors

Electronic properties together with biocompatibility enables QDs to be used in electrochemical sensing.

8.5.1 Uric Acid Biosensor

Zhang et al. (2006) performed UA ($C_5H_4N_4O_3$) biosensing with uricase-ZnS QD/L-Cyst ($HSCH_2CH_2NH_2$) assembly; uricase is a peroxisomal liver enzyme that catalyzes the oxidation of UA to allantoin during purine catabolism. Covalent binding of uricase and L-Cyst occurs through the free carboxyl groups on the surface of ZnS QDs. Properties of carboxyl group functionalized ZnS QDs such as solubility, conductivity, and more binding sites for increased enzyme loading imparted high electrocatalytic performance to the sensor.

8.5.2 Hydrogen Peroxide Biosensor

Where are H_2O_2 measurements required? The determination of hydrogen peroxide content is of great importance in many fields, including industrial environmental protection and clinical control.

HRP is one of the heme-containing redox enzymes with molecular weight of approximately 42,000 Da. The electron transfer converts Fe(III) in the heme of HRP to Fe(II). This process catalyzes some chemical reactions, such as the reduction of H_2O_2. Unfortunately, strong adsorption of HRP on the electrode surface causes denaturation. On the other hand, the electrochemically active centers in HRP are always buried deeply in its extended three-dimensional structure, which makes direct electron transfer between HRP and the electrode surface very difficult.

Wang et al (2009) immobilized HRP with lipophilic (fat-liking: the ability of a chemical compound to dissolve in fats, oils, lipids, and nonpolar solvents) CdSe/ZnS QDs on a GCE surface. It was found that HRP transfers electron directly onto the GC electrode only when the electrode was modified with QDs through vapor deposition. This may be due to the slow vapor deposition process of QDs, which results in good combination of HRP and QDs, and HRP adjusts its molecule to an appropriate orientation for direct electron transfer. Based on the direct electrochemistry of HRP, they fabricated an H_2O_2 biosensor. It was also found that the modified electrode could be used as a sensor for H_2O_2, and the linear range of detection was 5.0×10^{-6}–1.0×10^{-4} M, with a detection limit of 2.84×10^{-7} M. The sensor exhibited reproducibility and stability.

Compared with HRP, which is often used to construct hydrogen peroxide biosensors, Hb has many advantages, such as commercial availability, moderate cost, and its known and documented structure besides its intrinsic peroxidase activity; therefore, the use of Hb for the construction of hydrogen peroxide biosensors is possible.

Hb-on-nano cadmium sulfide (CdS)-modified electrode is used as an H_2O_2 biosensor. The forbidden bandwidth of nano CdS is 2.42 eV at the room temperature, and its photovoltaic effect can be induced by the visible light, generating the oxidative cavities and reductive electrons. The electrons can transfer to the surface of particles rapidly because of the nanoscale effect of

nano CdS, resulting in a higher charge detaching efficiency resulting in the enhanced electron transfer reactivity and catalytic activity. Hb is a typical heme protein. However, direct electron transfer between Hb and electrode is not easy to achieve. So various nanomaterials have been employed to obtain the satisfactory electrochemical response of the protein.

Zhou et al. (2005) immobilized Hb with CdS NPs on pyrolytic graphite (PG) electrode. Cyclic voltammogram of Hb/CdS NPs co-modified PG electrode in 0.1 M PBS with pH 6.0 showed a pair of pronounced redox peaks ascribed to Hb, with the anodic peak at –239 mV and the cathodic peak at –337 mV (versus SCE, saturated calomel electrode), respectively. The peak separation is about 98 mV. Hb cannot exhibit electrochemical response without the help of CdS NPs. If Hb alone is modified on the PG electrode surface, no wave can be observed. Even if the protein is immobilized onto the surface of CdS NPs, instead of being entrapped in the NPs, fine redox waves still cannot be obtained.

Linear relationship between the reductive peak current and the H_2O_2 concentration was obtained from 5.0×10^{-6} to 4.0×10^{-4} mol L^{-1}, on the basis of which an H_2O_2 biosensor might be developed.

Xu et al. (2007) proposed a hydrogen peroxide biosensor based on the direct electrochemistry of hemoglobin modified with quantum dots. They described the direct electrochemical behavior of Hb modified with QDs (CdS) at a normal GP electrode. Figure 8.11 shows the modification of Hb with QDs. Because of chemical interaction between CdS and Hb, the modified electrode is more stable. Hb modified with CdS showed direct electron transfer and the modified electrode was sensitive to hydrogen peroxide. Solution of Hb modified with CdS was cast onto the surface of the GP electrode and dried in air. Then, Nafion was used to cover the Hb–CdS film as a binder to hold the film stably on the electrode surface. The Nafion/CdS–Hb/GP electrode can be used as a hydrogen peroxide biosensor because of its good peroxidase-like activity bioactivity. This biosensor showed an excellent response to the reduction of H_2O_2 without the aid of an electron mediator. The catalytic current showed a linear dependence on the concentration of H_2O_2 in the range 5×10^{-7}–3×10^{-4} M with a detection limit of 6×10^{-8} M.

8.5.3 CdS Nanoparticles Modified Electrode for Glucose Detection

Huang et al. (2005) studied the direct electrochemistry of glucose oxidase (GOD) adsorbed on a CdS NPs modified PG electrode. PG is a unique form of GP manufactured by decomposition of a hydrocarbon gas at very high temperature in a vacuum furnace resulting in an ultrapure product which is near theoretical density and extremely anisotropic.

The enzyme showed significantly enhanced ET reactivity. GOD adsorbed on CdS NPs maintained its bioactivity and structure, and could electrocatalyze the reduction of dissolved oxygen, which resulted in a great increase of the reduction peak current. Upon the addition of glucose, the reduction peak current decreased, which could be used for glucose detection. CdS NPs

$$Cd^{2+}{}_n\text{–Hb}+ nS^{2-} \Leftrightarrow (CdS)_n\text{–Hb}$$

$$nCd^{2+}+Hb \Leftrightarrow Cd^{2+}{}_n\text{–Hb}$$
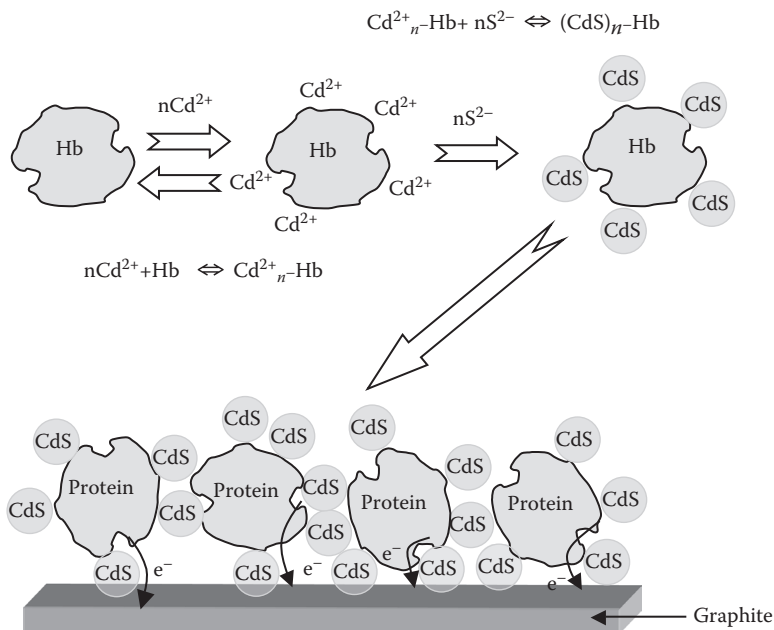
FIGURE 8.11
Process of preparing CdS–hemoglobin (Hb)/graphite electrode. (After Xu, Y. et al., *J. Biol. Inorg. Chem.*, 12, 421, doi: 10.1007/s00775-006-0198-2, 2007.)

provided an ideal environment for retaining the enzyme activity as well as promoting the ET reactivity of GOD. A mediator-free glucose biosensor was prepared. Considering its good stability and selectivity along with low cost, this biosensor showed great promise for the rapid determination of glucose.

8.5.4 QD Light-Triggered Glucose Detection

Schubert et al. (2009) demonstrated the applicability of quantum dots for NADH sensing. NADH is the abbreviation for reduced form of NAD (nicotinamide adenine dinucleotide), i.e., nicotinamide adenine dinucleotide: hydrogen reduced. It is a coenzyme that incorporates niacin and is involved in the Krebs cycle. NADH sensing is based on the immobilization of CdSe/ZnS nanocrystals on gold electrode. The use of quantum dots on the electrode provides a photoswitchable interlayer allowing the spatial readout of the sensor surface. Photogenerated excitons of semiconducting nanocrystals can electrically communicate with electrode surfaces and lead to an anodic or cathodic photocurrent. The current signal can be triggered by illumination of the quantum dot-modified electrode surface. Because of photoexcitation, electron hole pairs are generated in the quantum dots. Excited conduction-band electrons of the quantum dots can be transferred to an electrode or to

an electron acceptor in solution. Electrons can also be transferred from an electrode or a solubilized electron donor to valence-band holes in the quantum dots. Therefore, a quantum dot layer between the electrode and a redox system can be used for a light-triggered readout of the electron transfer reaction with the electrode.

The aforementioned light-induced carrier generation provides the basis for the combination of the CdSe/ZnS electrode with a NADH-producing enzyme reaction for the light-triggered detection of the corresponding substrate. Schubert et al. (2009) showed that glucose detection is possible with such an electrode system by photocurrent measurements. To prepare the electrodes (Figure 8.12), the quantum dots were first modified with a dithiol, 1,4-benzenedithiol (BDT), $C_6H_6S_2$ and afterward immobilized on the gold electrodes (Au-[QD-BDT]). A thiol is an organosulfur compound that contains a carbon-bonded sulfhydryl (–C–SH or R–SH) group (where R represents an alkane, alkene, or other carbon-containing moiety); a dithiol is a compound having two thiol groups. For immobilization of the CdSe/ZnS nanocrystals, their capping ligand, trioctylphoshine oxide (TOPO), $C_{24}H_{51}OP$, from synthesis was exchanged by BDT. TOPO is a tertiary alkylphosphine that can be used as an extraction or stabilizing agent. The use of a small dithiol provided the possibility of replacement of the original ligand TOPO, in a first step as well as the strong coupling of the nanocrystals to the gold electrode surface via chemisorption in a second step (Au-[QD-BDT]). The bare gold electrode showed no photocurrent, but after quantum dot immobilization a negative photocurrent in the range of 10 nA was observed at an applied potential of +50 mV versus Ag/AgCl, 1 M KCl. The detection of NADH was possible in the range from 20 μM to 2 mM. The detection of glucose using glucose dehydrogenase was effectively demonstrated.
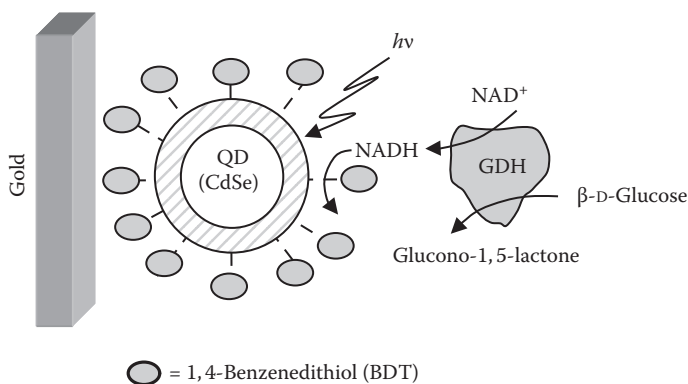
FIGURE 8.12
Glucose detection due to the catalytic production of NADH by the enzyme glucose dehydrogenase (GDH) in solution, performed at CdSe/ZnS quantum-dot-modified gold electrodes with QDs immobilized on gold via the ligand 1,4-benzenedithiol (Au-[QD-BDT]). (After Schubert, K. et al., *Langmuir*, 26, 1395, doi: 10.1021/la902499e, 2009.)

8.6 Nanotube- and Nanowire-Based FET Nanobiosensors

Can a planar field effect transistor (FET) be used for detecting biochemicals? Yes, its capability for this purpose is long known (Agarwal et al. 2010). *Then what has limited its application?* Clearly, its lower sensitivity. Compared to the surface region of a planar device, due to the ultrahigh surface-to-volume ratio, carbon nanotubes (CNTs) and silicon nanowires (NWs) are ideal choices for sensors due to the introduced depletion/accumulation of charges near the surface because of surface binding/adsorption of foreign molecules and species (Figure 8.13). The physical properties restraining sensor devices fabricated in planar semiconductors are readily triumphed over by exploiting nanoscale FETs, and the reasons are: (i) Binding to the surface of a carbon nanotube (CNT) NW leads to depletion or accumulation of carriers in the "bulk" of the nanometer diameter structure (versus only the surface region of a planar device) and increases sensitivity to the extent that single-molecule detection is possible. (ii) The small size of CNT and NW building blocks and topical advances in assembly suggest that dense arrays of sensors could be prepared.

8.6.1 Nanotube versus Nanowire

CNTs could be considered an ideal material for sensing applications because: (i) Every atom in an SWCNT is located on the surface, leading to extreme sensitivity to the contiguous environment. (ii) Surface-modified CNTs are compatible (nontoxic) with living organisms such as cells, thus providing the appropriate interface between biological entities and electronic circuits. (iii) Also, nanotubes are readily synthesized from inexpensive precursors such as methane, or purchased from commercial sources with high purity.

Several properties of CNTs limit their development as nanosensors: (i) Existing synthetic methods produce mixtures of metallic and semiconducting CNTs, which make systematic studies intricate because metallic devices will not function as expected. The need to separate semiconducting nanotubes from metallic nanotubes and a nonuniform distribution of bandgaps leads to an inability to fine-tune the electronic properties. (ii) Flexible methods for the modification of CNT surfaces, which are required to prepare interfaces selective for binding a wide range of analytes, are not well established. Long-term, water-stable modification with bioreceptor molecules for specific and sensitive biosensors has proved to be a challenge.

NWs of semiconductors such as Si do not have these limitations, as: (i) They are always semiconducting, and the dopant type and concentration can be controlled, which enables the sensitivity to be tuned in the absence of an external gate. (ii) In addition, it should be possible to exploit the giant
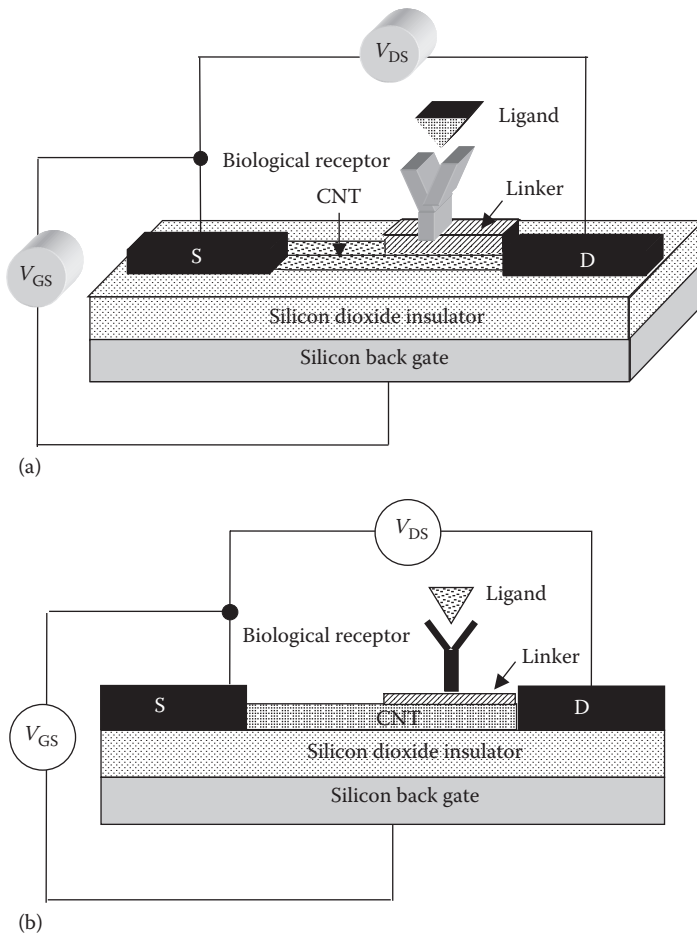
FIGURE 8.13
Biosensor based on carbon nanotube field-effect transistor: (a) three-dimensional view and (b) cross-section. The electrical current through the channel CNT is altered upon binding the ligand–receptor (because of the charge of the ligand) and is detected, by the change of drain–source current.

knowledge that exists for the chemical modification of oxide surfaces, e.g., from studies of silica and planar chemical sensors.

8.6.2 Functionalization of Si NWs

Two approaches for functionalizing the Si NWs are element-doping and surface chemical modification (Yang et al. 2006). Compared with the element-doping method, chemical modification is easily controlled and able to anchor various functional groups on the surface of Si NWs. For Si NWs, the linker molecule of choice depends on whether or not the wire has an oxide

coating. Accordingly, the appropriate functionalization methods are briefly described as under:

1. *Functionalization of Si NWs coated with a native oxide layer:* A variety of linker molecules have been designed to bind to the native oxide coating the Si NWs (Curreli et al. 2008). Among these, alkoxysilane derivatives (any alkoxy derivative of a silane) are the most widely used linkers. The Si-methoxide, CH_3SiO, or Si-ethoxide, C_2H_5OSi, reacts with the surface OH group, anchoring the linker molecule to the silicon oxide surface. Common linkers for Si/SiO_2 NW functionalization are 3-(trimethoxy-silyl)propyl aldehyde (TMSPA), and Aminopropyltrimethoxysilane (APTMS). Another popular linker molecule is 3-aminopropyl-triethoxysilane (APTES), $C_9H_{23}NO_3Si$. This reagent yields a surface coated with amino groups. These $-NH_2$ groups are activated toward bioconjugation (the process of coupling two biomolecules together in a covalent linkage) by using the proper coupling reagent.

2. *Functionalization of H-terminated Si NW surfaces:* The silicon oxide coatings are easily etched away by submerging the NWs into dilute HF. This replaces the native oxide layer with a hydride-terminated silicon surface. This hydride-terminated surface is found to be air-stable for several days. The Si–H bond is rapidly photodissociated with ultraviolet (UV) radiation to generate radical species on the Si surface. These radicals subsequently react with terminal olefin groups on linker molecules, thus forming stable Si–C bonds at the Si surface. An olefin, olefine, or alkene, is an unsaturated chemical compound containing at least one carbon-to-carbon double bond. This photochemical hydrosilylation treatment selectively functionalizes the Si NWs, but does not react with the underlying SiO_2 layer of the substrate. This photohydrosilylation treatment carried out using an olefin derivative of an easily cleavable carbamate (organic compound derived from carbamic acid (NH_2COOH)), followed by deprotection, results in Si NWs coated with amino groups. The $-NH_2$ groups are then used to physically adsorb probe single-stranded (ssDNA) on the Si NWs, and attach several biotin derivatives, antibodies, and probe ssPNA.

Photohydrosilylation is the photochemically induced addition of a hydrosilane across the double bond of an alkene or alkyne giving a more substituted silane. It is usually applied to cases where the hydrosilane is part of a silicon surface. The reaction uses UV irradiation and is initiated by an added photoactive compound such as a platinum catalyst.

8.6.3 DNA and Protein Detection

If the bound analyte molecule carries a charge opposite to the main carriers in the FET, then charge carriers accumulate under the bound analyte, thus

causing an increase in the device conductivity. This mechanism is shown in Figure 8.14a, where a negatively charged molecule such as DNA binds to the P-type NW, causing a gathering of hole carriers, thus resulting in an increase in conductivity. In contrast, analytes with molecular charges same as that of the main carriers in the FET lead to depletion of main carriers beneath the bound analyte, causing a decrease in conductivity. The latter case is shown in Figure 8.14b, where a positively charged molecule, such as a protein below its isoelectric point (pI), depletes the carriers upon binding to the NWs. Isoelectric point is the pH at which a particular molecule or surface carries no net electrical charge.

A traditional approach to DNA detection utilizes a ssDNA probe (attached to the NW) that hybridizes with its complementary single strand. This strategy requires a minimum 10 mM of electrolyte concentration at room temperature to ensure robust hybridization. Operating at such high salt concentrations could significantly diminish the device sensitivity. Therefore, using a different probe/capture molecule such as the uncharged PNAs (peptide nucleic acids) offers several advantages over DNA probes. PNA–DNA hybridize at low electrolyte concentrations, conditions under which there is a long Debye length (the length in which mobile charge carriers screen out the external electric field) and the devices are expected to be particularly sensitive. It was found that immobilization of only the probe DNA produced a resistance increase of greater than 300%, whereas for the PNA probe, the resistance of the device was unchanged. Upon hybridization with complementary DNA, a 14% conductance change was observed for the DNA probe in contrary to a 200% increase in resistance for the PNA-modified device.

Cancer-derived proteins or existing proteins present at abnormal concentration are biomarkers that are tracked to monitor the progress of cancer. Therefore, devices that quantify the level of biomarkers in serum or other human samples have impending applications in the diagnosis of cancer or other diseases.

The first example of a real-time, electrical detection of a protein from a solution using a FET nanosensor, reported by Cui et al. (2001), used the biotin–SA probe-receptor complex. The interaction of biotin and SA was monitored using P-type Si NWs, and binding was observed with a limit of detection (LOD) of 10 pM of SA, at $\lambda_D \sim 3.4$ nm. The conductance of the NW device increased rapidly in compliance with the gating effect of the negatively charged SA. Upon addition of pure buffer, the signal remained unwavering because of the high binding affinity between biotin and SA. A system with lower binding affinity than biotin–SA, biotin–antibiotin (monoclonal Ab), was then preferred to prove reversible binding and regeneration of the sensing surface. As expected, the flow of fresh buffer rapidly detached the bound antibiotin and the device conductivity returned to baseline, suggestive of successful recovery of the sensor.
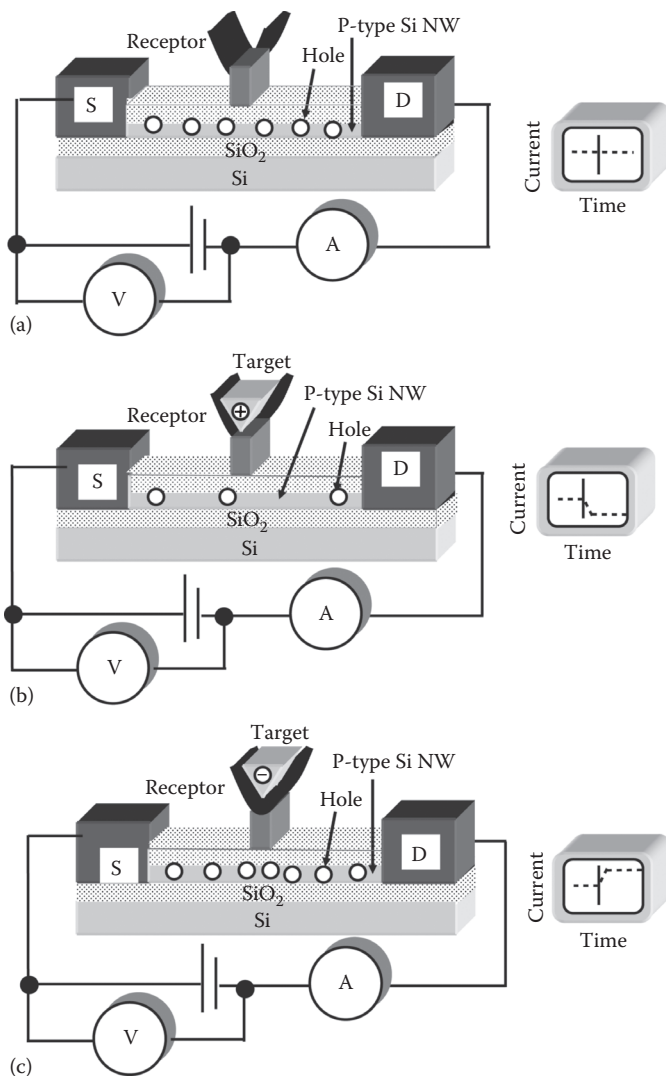
FIGURE 8.14
P-type oxidized silicon nanowire: (a) Without any target molecule bound to receptor site. (b) With a positively charged target molecule bound to the receptor site causing hole depletion in NW and decrease in its conductance. (c) With a negatively charged target molecule bound to the receptor site causing hole accumulation in NW and increase in its conductance.

8.7 Cantilever-Based Nanobiosensors

Do these cantilevers have any relation to atomic force microscopy (AFM) cantilevers? Yes, microcantilevers are derived from the microfabricated cantilevers used in AFM and are based on the bending induced in the cantilever when a biomolecular interaction takes place on one of its surfaces. They translate the molecular recognition of biomolecules into nanomechanical motion (from a few nm to hundreds of nm). They are typically made of silicon/silicon nitride or polymer materials, with dimensions ranging from tens to hundred of μm long, some tens of μm wide, and hundreds of nm thick (Carrascosa et al. 2006). Moreover, these devices can be fabricated in arrays comprising ten to thousands of microcantilevers. Hence, they are promising alternatives to current DNA and protein chips because they permit parallel, fast, real-time monitoring of thousands of analytes (e.g., proteins, pathogens, and DNA strands) without any need for labeling.

What are the capabilities of nanocantilevers? When fabricated at the nanoscale (nanocantilevers), the expected LODs are in the femtomole (fmol = 10^{-15} mol) to attomole (amol = 10^{-18} mol) range with the astonishing possibility of detection at the single-molecule level in real time.

Microcantilevers themselves do not have any inherent chemical selectivity. Then *how is an ordinary solid-state surface transformed into an intelligent sensor surface, recognizing and identifying complex biological systems?* The key to using microcantilevers for selective detection of molecules is the ability to functionalize one surface of the silicon microcantilever in such a way that a given target molecule will be preferentially bound to that surface upon its exposure. The sensitivity of detection is greatly enhanced by applying an appropriate coating to one cantilever surface. Selective biochemical sensors are prepared by coating or covalently binding a molecular recognition agent, i.e., a molecule or a polymer that has a strong and specific biomolecular interaction with a guest molecule, to the microcantilever surface (Hansen and Thundat 2005).

This strategy allows microcantilever sensors to measure extremely small changes due to molecular adsorption and, essentially for that reason, they are extremely sensitive biosensors; with the cantilever technique, it is possible to detect surface stress as small as about 10^{-4} N m^{-1}. Such measurement is also quantitative; it is related to the concentration of the analyte being detected. Nonetheless, it must be clearly borne in mind that the factors and the phenomena responsible for the surface-stress response during molecular recognition still remain ambiguous and confusing for biological adsorption, due to the complexity of the interactions involved.

What are the two main approaches for using cantilevers in biosensing?

1. By immobilizing a ligand on one side of the cantilever and placing it in contact with a receptor in solution, the cantilever bends in response to a change in surface stress generated by ligand–receptor

binding; the greater the binding energy, the greater the bending. A ligand is an ion or molecule that binds to a central metal atom to form a coordination complex. The bending causes movement of the cantilever tip of the order of 1–100 nm.

2. Also, the resonance frequency of a microcantilever varies sensitively with molecular adsorption. In the dynamic or resonance mode, cantilevers are excited close to their resonance frequency, which is typically in the kHz or even MHz range. When an additional mass is attached to the oscillating cantilever, its resonance frequency changes (adding a mass lowers the resonance frequency). This is not surprising since at a first approximation, cantilevers behave like a harmonic oscillator, an ideal oscillating spring-mass system.

What is the explanation for bending and deflection of the cantilever? What are compressive and tensile stresses? Changes in the environment around or directly on the surface of the cantilevers create a mechanical stress in the surface, which leads to an expansion or contraction of the cantilever surface. If this stress acts only on one side of the cantilever, then the asymmetrically stressed structure bends and the cantilever is deflected. For simple isotropic materials, the surface stress relates the reversible work dW of deforming a surface (elastically) to the change in surface area dA by

$$dW = \sigma \times dA \qquad (8.1)$$

For spontaneous processes, i.e., $dW < 0$, and assuming a positive stress ($r > 0$), dA has to be negative, the surface desires to contract and the stress is said to be tensile. For a negative surface stress ($r < 0$), dA has to be positive, the surface desires to expand and the stress is said to be compressive. In a nutshell, expansion of a surface is defined as compressive surface stress and contraction as a tensile surface stress.

Two general comments must be made here. Firstly, compressive and tensile stresses have always to be associated with one specific surface of a cantilever since an overall upward bending can be caused by either a tensile stress at the top surface or by a compressive stress at the bottom surface. Furthermore, it has to be emphasized that when discussing about surface stresses detected by cantilever sensors, one always refers to a change in surface stress, but not to an absolute stress.

Why do scientists recommend using cantilever arrays for measurements instead of a single cantilever? Independent of the different modes it is always advantageous to use several cantilevers in parallel. It is evident that one can get more information out of a single experiment when using more sensors in parallel. A more sophisticated argument for parallel sensing lies in the way biologists deal with the fact that their systems under investigation (biomolecules and cells) are not as general and ideal as typical physical systems (e.g., atoms and crystals). *How do biological systems differ from physical systems?* Biological

systems are normally more complex and more unique than physical systems and their properties depend much more strongly on their history and actual environment. Therefore, well thought-out control experiments constitute an especially important part of any experiment in the life sciences.

Cantilevers are temperature-sensitive sensors and might also respond to changes in buffer composition. In addition, several other molecules in the sample and not only the target molecules might interact with the sensor. One therefore needs a strategy to subtract these unwanted background signals from the "real" signal. In particular, first-time users make haste to obtain quick results with their samples, but are often not aware of the fact that signals from the environment in which the target molecules are presented can be much stronger than signals from the target molecules themselves. Therefore, researchers use arrays of cantilevers with several sensors in parallel: all cantilevers are physically identical and only differ in their surface coating, as shown for two cantilevers in Figure 8.15 to illustrate the concept. Their physical uniformity can be checked before an experiment by either applying a well-defined heat pulse and recording the thermal responses of the cantilevers or by measuring their individual resonance frequencies which should be identical to within at least about 5%. Some cantilevers are then made sensitive for the target molecules whereas others act as a reference for unspecific binding or for physical signals such as temperature, refractive index changes, or different buffers. Experiments with complex biological samples are nowadays performed with cantilever arrays of up to eight cantilevers in parallel.

In contrast to competing technologies, such as surface plasmon resonance (SPR), which measures changes in dielectric constant, cantilever arrays are
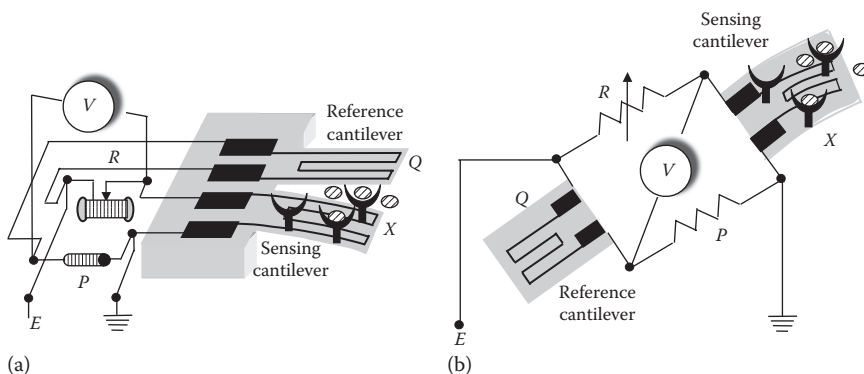
FIGURE 8.15
Measurement of microcantilever deflection using meandering piezoelectric resistors: (a) Placement of sensing and reference cantilevers and (b) circuit drawn in common Wheatstone bridge configuration for easy understanding. Bending of the cantilever causes elongation of the resistor on its surface changing its resistivity. The inert reference cantilever provides differential measurements and helps in cancelling out the parasitic effects of temperature and chemical interactions.

not limited by mass sensitivity but operate via an entirely different principle, detecting tiny changes in in-plane forces, called surface stress.

8.7.1 Biofunctionalization of the Microcantilever Surface

Coating or functionalizing a sensor is a critical preparation step because the recognition layer defines the application and performance of a sensor. To modify the cantilevers with chemically selective layers, surface chemistry and molecular recognition methods developed for other sensors are utilized, especially those for quartz crystal microbalance (QCM), surface acoustic wave devices (SAW), chemical field-effect transistors (chemFET), and surface-enhanced Raman spectroscopy (SERS) sensors.

   Many approaches can be used to immobilize the molecular recognition agents to the microcantilever sensor, depending upon the final application. *What are the main prerequisites of the immobilization process?* The immobilization process should: (i) Avoid any change in the mechanical properties of the cantilever. (ii) It should be uniform, in order to generate a surface stress as large as possible. (iii) It should allow accessibility by the target molecule. (iv) Biomolecules should be tightly attached to the sensor surface but still flexible and functional as in their natural environment. (v) Surface coatings have to be reliable, they should be robust against changes in buffer and temperature and ideally withstand repetitive detection and cleaning cycles. Especially for cantilever sensors, interactions on top of the sensing layer should be fully transferred to the underlying substrate favoring a dense and covalent surface functionalization with receptor molecules close to the surface.

   Usually, cantilevers already show two distinct surfaces, e.g., a gold top surface coated with a thin layer of gold (20–100 nm) and silicon dioxide bottom surface. Depending on the final application for the molecular recognition assay, either the silicon dioxide side or the gold side of the cantilever can be used. *Are the coating methods different for the two surfaces?* Silane chemistry is employed for silicon dioxide and SAM thiol chemistry for gold-coated cantilevers.

1. The silicon dioxide surface is modified by amino- or mercapto-silane monolayers whose end groups are further cross-linked to receptor molecules. In addition, highly positively charged molecules are electrostatically bound to the negatively charged silicon dioxide. Silane is a molecule containing a central silicon atom bonded to two types of groups: alkoxy groups and organo-functional groups. These two types of groups exhibit different reactivity and allow sequential reactions. Aminopropyltrimethoxysilane has the formula $H_2N(CH_2)_3Si(OCH_3)_3$. Formula of 3-mercaptopropyltrimethoxysilane is $C_6H_{16}O_3SSi$.

   Silanes are useful compounds for anchoring an organic layer to an inorganic substrate. They contain Si–O bonds that react with surface hydroxides and pendant hydrocarbon chains to link with

the organic overlayer, such as a polymer or protein molecules. The chemistry of the pendant group is alterable by selective chemical reactions. Silanation therefore gives us a technique to tailor the adhesion properties of a surface or to change its biocompatibility. One method of forming thin, uniform layers of silanes is to take advantage of a "self-assembly" process. This occurs primarily in long-chain hydrocarbon molecules adsorbed on surfaces. The molecules associate into well-ordered domains after they adsorb. This association leads to the formation of well-organized, SAMs, called SAMs, on surfaces. A SAM is a two-dimensional film, one molecule thick, covalently assembled at an interface.

2. SAMs are formed by exposing the solid substrates to amphiphilic molecules (molecules having a polar, water-soluble group attached to a nonpolar, water-insoluble hydrocarbon chain) with chemical groups having a strong affinity for the substrate. Alkanethiols absorb spontaneously from solutions on the Au surface. Gold is the most frequently used metal because it does not form a stable oxide layer under ambient conditions.

On the other hand, because of the high affinity of sulfur groups for gold, the gold surface is modified by thiolabeled nucleic acids or proteins exposing cysteines (α-amino acids with the chemical formula $C_3H_7NO_2S$) showing sulfur atoms at their surface. Thiolated poly(ethylene glycol)s (SH-PEG), act as inert layers, preventing molecular adsorption; molecular formula of PEG is $C_{2n+2}H_{4n+6}O_{n+2}$.

For organosilane modification and thiol- SAMs, *dip coating* is the preferred method for functionalization to allow for high-density immobilization on the cantilever surface. All reactive surfaces of the cantilever and substrate that are exposed to the modifying solution (s) acquire a coating. Organosilane (organic derivative of a silane containing at least one carbon to silicon bond) coatings are of the order of a monolayer thick but become multilayer coatings upon extended exposure to the solution. But thiol SAMs are self-limited to coverages of a monolayer or less of the thiol on a gold film.

Immobilization of both DNA and proteins on gold surfaces using thiol-SAMs is a well-known chemistry and has been widely practiced in many other biosensing applications. Direct coupling of DNA probes by self-assembly of thiol-labeled oligonucleotides is a common, easy technique for gold-coated microcantilevers. Herne and Tarlov (1997) characterized the immobilization of single-strand DNA oligonucleotides on gold via sulfur linkage. The sulfur atom causes a great change in surface stress during DNA immobilization in contrast to non-modified DNA. Most DNA-biosensing applications performed with microcantilever technology are based on this strategy.

How are proteins adsorbed on gold surfaces of cantilevers? The covalent adsorption of proteins on gold surfaces of cantilevers is achieved by a wide variety of chemical procedures, ensuring the reproducibility and the stability of the protein coating. A very interesting one is covalently immobilization of carboxylate-terminated alkanethiols (e.g., 11-MUDA, $C_{11}H_{22}O_2S$) followed by esterification (the chemical process for making esters, compounds of the chemical structure R–COOR', where R and R' are either alkyl or aryl groups) of the carboxylic groups with EDC, $C_8H_{17}N_3$, and NHS, $C_4H_5NO_3$. Carboxylate is any salt or ester of a carboxylic acid (R–COOH, where R is some monovalent functional group) having a formula of the type $M(RCOO)_x$, where M is a metal and R an organic group. Figure 8.16 shows the deposition of mercapto-derivatized linking molecule on the gold cantilever surface; "mercapto" means containing the univalent radical –SH: *mercaptopurine*. Figure 8.17 illustrates how the mercapto-derivatized linking molecule forms an amide bond on treatment with EDC and NHS. The reaction is used for the immobilization of the protein on the gold-coated cantilever. Here, the intermediate active ester (the product of condensation of the carboxylic group and NHS) reacts with the amine (organic compounds that contain nitrogen and are basic) function of proteins to yield the amide bond and the final immobilization of protein. Amide is a chemical compound formed from an organic acid by the substitution of an amino (NH_2, NHR, or NR_2) group for the hydroxyl of a carboxyl (COOH) group. An amide bond is a chemical bond formed between two molecules when the carboxyl group of one molecule reacts with the amine group of the other molecule, thereby releasing a molecule of water (H_2O).

Another option is to use cystamine-modified with glutaraldehyde and the subsequent attachment of the protein through an amine group; cystamine is an organic disulfide. The immobilization of different bioreceptors on each cantilever of an array is a complicated task. There are many commercial platforms devoted to the specific functionalization of individual cantilevers.

(a) (b)
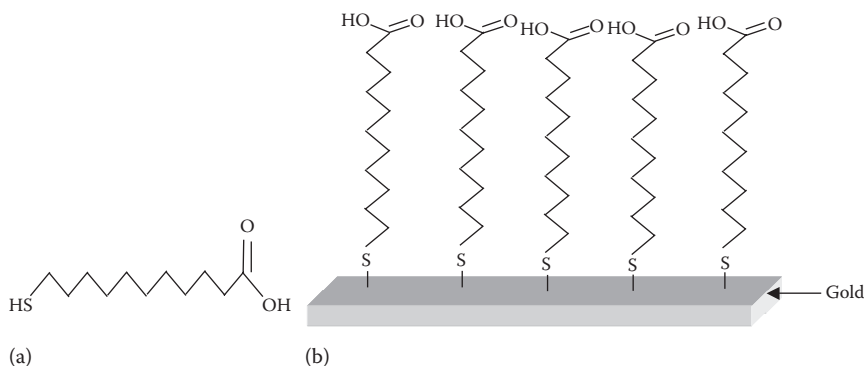
FIGURE 8.16
(a) 11-Mercaptoundecanoic acid. (b) SAM of mercaptoundecanoic acid on gold surface.
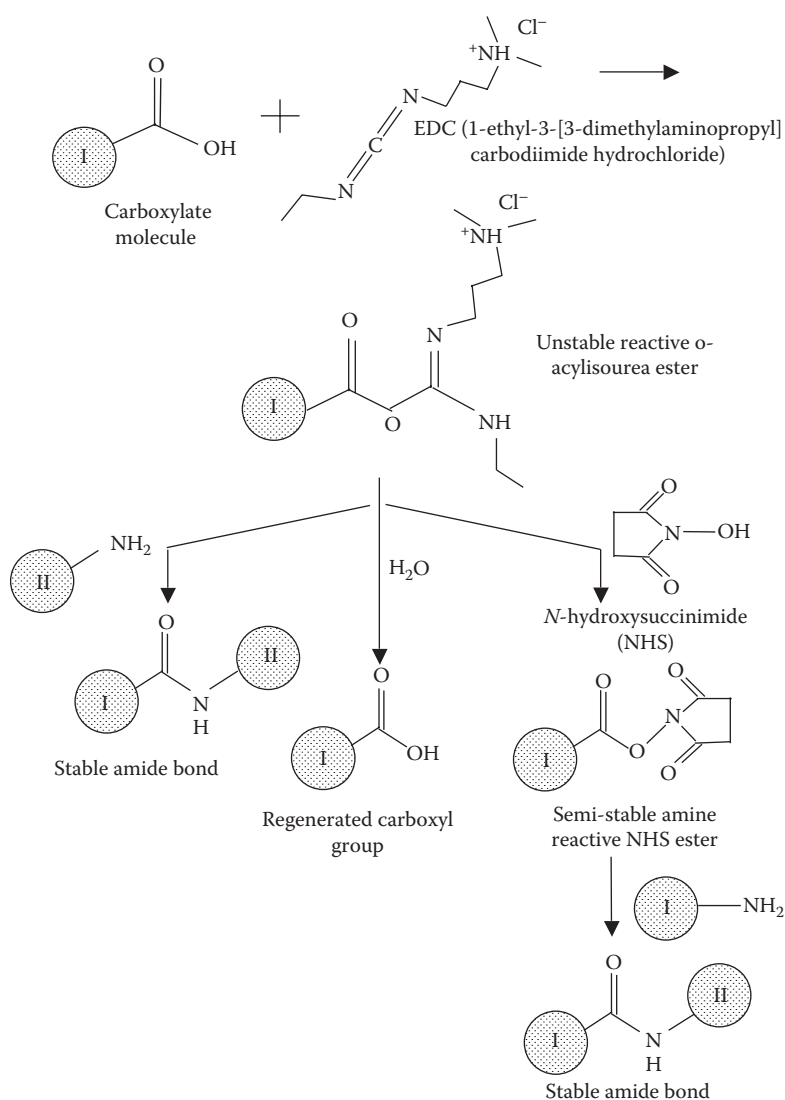
FIGURE 8.17
Esterification of the carboxylic groups with 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) and *N*-hydroxysuccinimide (NHS). EDC is a water-soluble derivative of carbodiimide that catalyzes the formation of amide bonds between carboxylic acids or phosphates and amines by activating carboxyl or phosphate to form an O-urea derivative. This derivative reacts readily with nucleophiles. NHS is often used to assist carbodiimide coupling in the presence of EDC. The reaction includes formation of the intermediate active ester (the product of condensation of the carboxylic group and NHS).
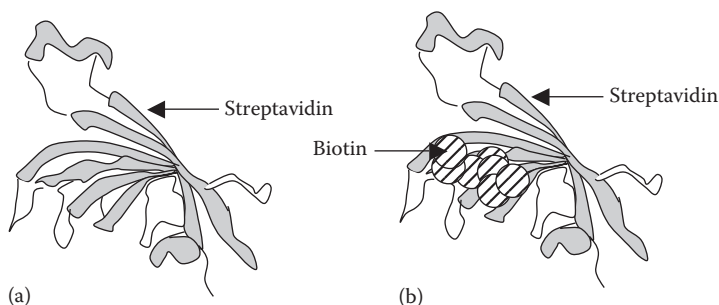
FIGURE 8.18
(a) Structure of streptavidin and (b) biotin molecules represented by spheres clinging to streptavidin.

Several approaches exist for exposing individual cantilevers of a cantilever array to specific biomolecules. The microfabricated cantilevers are incubated in individual glass capillaries or standard pipette tips filled with coating molecules. They are dipped in the channels of an open microfluidic network, or they are spotted with microliter drops of receptor molecules using an ink jet type of device. Special care is taken for providing a homogeneous coating at the hinge region, where the flexible cantilever beam is connected to the bulk silicon. Bending in this area influences the deflection of the free end of the cantilever much more strongly than a bending somewhere in the middle or close to the end of the cantilever.

Biotin–streptavidin interactions are well-studied binding partners that interact with very high affinity (Figure 8.18). Shu et al. (2007) investigated biotin–streptavidin binding interactions using microcantilever sensors. Three structurally different biotin-modified cantilever surfaces were produced as shown in Figure 8.19. The cantilever response to the binding of SA on these biotin-sensing monolayers was compared. The mechanical response of the cantilever strongly depends upon the nature of the biotin-modified surfaces: (i) Biotin-PEG-coated microcantilevers did not bend upon the injection of SA. (ii) Biotin-HPDP, (*N*-(6-(Biotinamido)hexyl)-3′-(2′-pyridyldithio)-propionamide (a reversible biotinylation reagent)-coated microcantilevers bent downward. (iii) Biotin-SS-NHS (succinimidyl 2-(biotinamido)-ethyl-1,3′-dithiopropionate)-coated microcantilevers bent upward. Biotin-SS-NHS enables simple and efficient biotinylation of antibodies, proteins, and other primary amine-containing molecules.

8.7.2 Biosensing Applications

The first applications of cantilever sensors for biological systems were reported in 1996 with single cantilevers (Butt 1996). The first biosensing experiments with cantilever arrays were demonstrated in 2000 showing the proof of principle for DNA detection and the ability to identify single-base
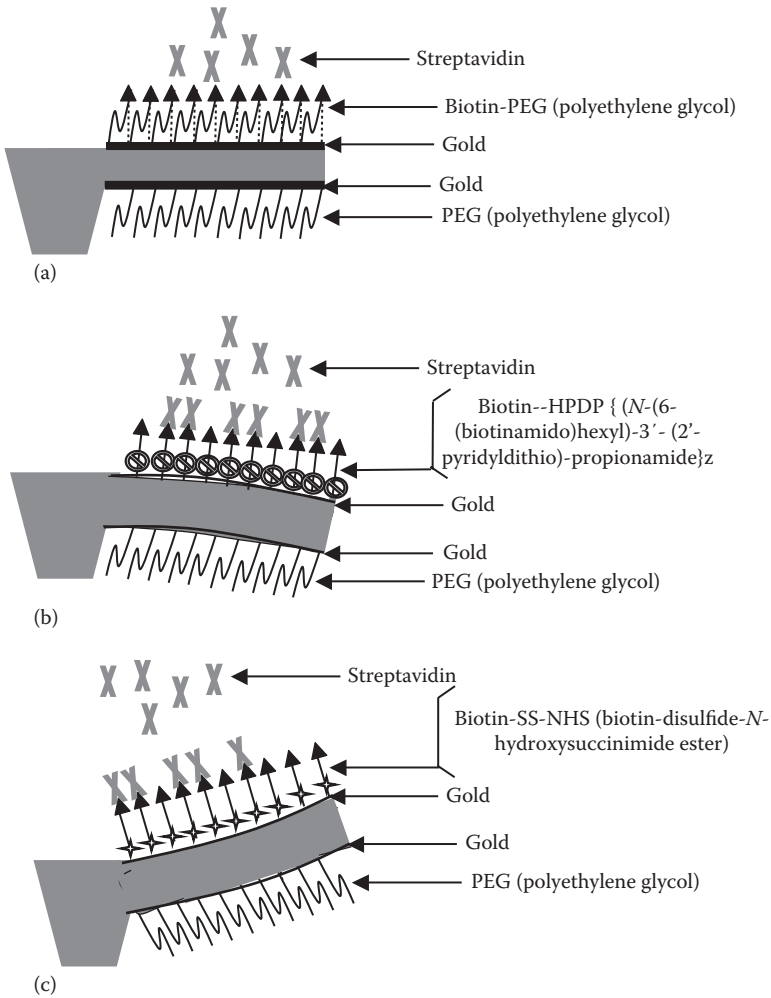
FIGURE 8.19
Difference in behavior of microcantilevers with various surface modifications on top gold surfaces: (a) biotin-PEG; (b) biotin-HPDP; and (c) biotin-SS-NHS. Bottom gold surfaces of all the microcantilevers are coated with PEG known to hinder nonspecific adsorption of proteins. (After Shu, W. et al., *Biosens. Bioelectron.*, 22, 2003, 2007.)

mismatches between sensing and target DNA oligonucleotides. This excellent example in the field of genomics involved the detection of a single-base mismatch with an LOD of 10 nM reported by Fritz et al. (2000). For such detection, they used an array of two cantilevers with a control (noncomplementary) oligonucleotide in one of them and the DNA probe (complementary) immobilized in the other, giving hybridization-deflection signals as small as 10 and 16 nm for 12-mer and 16-mer DNA targets, respectively, and displaying a deflection noise of 0.5 nm. The length of the oligonucleotide

is usually denoted by "mer," e.g., a fragment of 12 bases would be called a 12-meroligonucleotide.

More detailed experiments include showing the dependence of cantilever bending on the length, grafting density, and orientation of DNA oligonucleotides on surfaces, and demonstrating the ability to detect different sequences in parallel and within a high background of unspecific sequences.

Wee et al. (2005) have reported the detection of PSA (prostate-specific Ag: a protein manufactured exclusively by the prostate gland) and C-reactive proteins (CRP: one of the plasma proteins known as acute-phase proteins, produced by the liver and found in the blood), which is a specific marker of cardiac disease, by means of an electromechanical biosensor using self-sensing piezoresistive microcantilevers.

In addition, a novel development for early osteosarcoma (malignant bone tumor common in children and young adults) detection has been described, sensing the interactions between vimentin (any of a group polypeptides that polymerize to form filaments in the cytoskeleton) antibodies and Ags with a single cantilever-based biosensor (Milburn et al. 2005) and supporting the idea that cantilever biosensors can provide a suitable platform for life-sciences research.

The most elaborate experiments with nucleic acids report the detection of non-amplified RNA (ribonucleic acid) in total RNA from a cell with a detection limit of 10 pM (Zhang et al. 2006). These experiments were done within a high background of unspecific molecules and showed for the lowest concentration a differential deflection of 10 nm, or a surface stress of 1 mN m^{-1} (milli newtons per meter).

Example 8.1

The length (l) of a silicon cantilever is 20 μm and its thickness (t) is 300 nm. Upon antigen–antibody coupling, its deflection (Δz) was 7 nm. Find its radius of curvature (R), and also the surface stress (σ) developed in the cantilever. Young's modulus of silicon $E = 1.69 \times 10^{11}$ N m^{-2}, and Poisson's ratio (μ) = 0.25.

The radius of curvature is calculated from the cantilever length l and the sensor deflection Δz from the formula

$$R^{-1} \approx \frac{3\Delta z}{2l^2} = \frac{3 \times 7 \times 10^{-9}}{2\left(20 \times 10^{-6}\right)^2} = 26.25 \, \text{m}^{-1} \tag{8.2}$$

Hence $R = 1/26.25 = 3.81 \times 10^{-2}$ m. The bending is related to the change of surface stress σ according to Stoney's formula given in Equation 4.42

$$\sigma = \frac{Et^2}{6R(1-\mu)} = \frac{1.69 \times 10^{11} \times \left(300 \times 10^{-9}\right)^2}{6 \times 3.81 \times 10^{-2} \times \left(1 - 0.25\right)} = 8.871 \times 10^{-2} \, \text{Nm} \tag{8.3}$$

Example 8.2

(a) A rectangular cantilever of silicon has dimensions: $40\,\mu m$ (length) $\times 5\,\mu m$ (width) $\times 0.7\,\mu m$ (thickness). In the dynamic mode, the cantilever was excited to its resonance frequency f_0. Determine this resonance frequency if the spring constant of the cantilever is $0.07\,N\,m^{-1}$, and density of silicon is $2329\,kg^{-3}$.

(b) Adsorption of biomolecules increased the mass of the cantilever by 100 pg and decreased the resonace frequency to f_1 by an amount $\Delta f = f_0 - f_1$. Find f_1 and Δf.

The resonant frequency f_0 of an oscillating cantilever is given by

$$f_0 = \left(\frac{1}{2\pi}\right)\sqrt{\frac{k}{m_{eff}}} \tag{8.4}$$

where
  k is the spring constant
  m_{eff} is the effective mass of the cantilever

The effective mass is used here because of the reason given in the following: All parts of the cantilever do not oscillate with the same amplitude. The largest deflection takes place near the free end with a decay to zero at the clamped end. Therefore, in the analytical model development, the cantilever is represented by a lumped mass model. The lumped mass m_{eff} is smaller than the real mass m, by a factor, which depends on the geometry of the cantilever. A good rule of thumb says that the effective mass is 1/3rd of the real mass. Gluing tips or mirrors on cantilevers adds their mass to the effective mass. In general, the effective mass is expressed as

$$m_{eff} = nm_b \tag{8.5}$$

where
  n is the factor $= 0.24$ for a rectangular cantilever
  m_b is the mass of the cantilever beam given by

$$m_b = l \times w \times t \times \rho \tag{8.6}$$

where
  l, w, t are respectively the length, width, and thickness of the cantilever
  ρ is the density of its material

Hence,

$$m_b = l \times w \times t \times \rho = 40 \times 10^{-6} \times 5 \times 10^{-6} \times 0.7 \times 10^{-6} \times 2329 = 3.261 \times 10^{-13}\ \text{kg} \tag{8.7}$$

from which $m_{eff} = 0.24 \times 3.261 \times 10^{-13}\,kg = 7.826 \times 10^{-14}\,kg$, and the resonance frequency is

$$f_0 = \left(\frac{1}{2\pi}\right)\sqrt{\frac{k}{m_{\text{eff}}}} = \frac{1}{2\times 3.14} \times \sqrt{\frac{0.07}{7.826\times 10^{-14}}} = 1.506\times 10^5 \text{ Hz} \qquad (8.8)$$

When adsorbates are uniformly deposited on the cantilever surface, the resultant mass change is

$$\Delta m = \frac{k}{4\pi^2 n}\left(\frac{1}{f_1^2} - \frac{1}{f_0^2}\right) \qquad (8.9)$$

where f_1 and f_0 are the resonance frequencies after cantilever loading and before loading. Equation 8.9 yields

$$100\times 10^{-12}\times 10^{-3} = \frac{0.07}{4\times(3.14)^2\times 0.24}\left(\frac{1}{f_1^2} - \frac{1}{\left(1.506\times 10^5\right)^2}\right) \qquad (8.10)$$

or,

$$1\times 10^{-13} = 7.395\times 10^{-3}\times\left(\frac{1}{f_1^2} - 4.4091\times 10^{-11}\right) \qquad (8.11)$$

or,

$$1\times 10^{-13} = \frac{7.395\times 10^{-3}}{f_1^2} - 3.261\times 10^{-13} \qquad (8.12)$$

or $f_1^2 = (7.395\times 10^{-3})/(4.261\times 10^{-13}) = 1.73551\times 10^{10}$. Therefore $f_1 = 1.31739\times 10^5$ Hz, and the frequency decreased by $(1.506–1.31739)\times 10^5 \text{ Hz} = 1.8861\times 10^4$ Hz.

8.8 Optical Nanobiosensors

As these have already been treated in the chapter on optical nanosensors, only few additional examples will be presented here to supplement what has been discussed before.

8.8.1 Aptamers

What are aptamers? Aptamers are single-stranded RNA or DNA sequences with ligand-binding ability, which have been developed to detect a variety

of molecular targets including small molecules, nucleic acids, proteins, and cancer cells. *How are aptamers created?* They are created *in vitro* through systematic evolution of ligands by exponential enrichment for the recognition of target analytes with high affinity and specificity. *What is the relevance of aptamers for biosensors?* They have become important tools for molecular diagnostics and therapeutics in the biosensing of disease-related proteins. The aptamer biosensors offer detection limits down to the nanomolar (10^{-9} mol) level.

How can aptamers be compared with antibodies for developing biosensors? These artificial nucleic acid ligands appear as attractive alternatives to natural receptors, such as antibodies. They possess several advantages for biosensor design compared to antibodies: (i) Due to their relative ease of isolation and modification, tailored binding affinity, and resistance against denaturation, specific aptamers can be selected *in vitro* for virtually any given target ranging from small molecules to large proteins and even cells, thus paving the way to developing a wide range of aptamer-based biosensors. (ii) Unlike antibodies or enzymes, DNA aptamers are usually chemically stable and readily commercially available at low cost. (iii) Many aptamers undergo significant conformational changes upon target binding, offering high flexibility in biosensor development.

8.8.2 Aptamer-Modified Au Nanoparticle as a Colorimetric Adenosine Nanosensor

For what purposes is adenosine sensing necessary? Adenosine is an endogenous (growing or developing from within, e.g., endogenous cholesterol is cholesterol that is made inside the body and is not in the diet) nucleoside (a structural subunit of nucleic acids: any of various compounds consisting of a sugar, usually ribose or deoxyribose, and a nitrogen base) interacting in various physiological processes which are also regulated by hormones and neurotransmitters (chemicals located and released in the brain to allow an impulse from one nerve cell to pass to another nerve cell). Adenosine modulates blood flow and neurotransmission and may be protective during pathological (of or concerned with diseases) conditions such as ischemia (restricted blood flow within tissues) and stroke. Adenosine triphosphate (ATP), having empirical formula $C_{10}H_{16}N_5O_{13}P_3$ and chemical formula $C_{10}H_8N_4O_2NH_2(OH)_2(PO_3H)_3H$, has important responsibilities in regulating cellular metabolism and biochemical pathways. ATP is a molecule found in any biological system to work as an energy carrier. Detecting ATP within cells can help observe energetic physiological processes, e.g., signal cascades, transport processes, etc. Signal cascade amplifies the initial signal caused by binding of ligand to receptor. ATP depletion is related to certain diseases, such as Parkinson's disease and ischemia (a restriction in blood supply). Adenosine sensor is needed for understanding its physiological actions and extent of receptor activation.
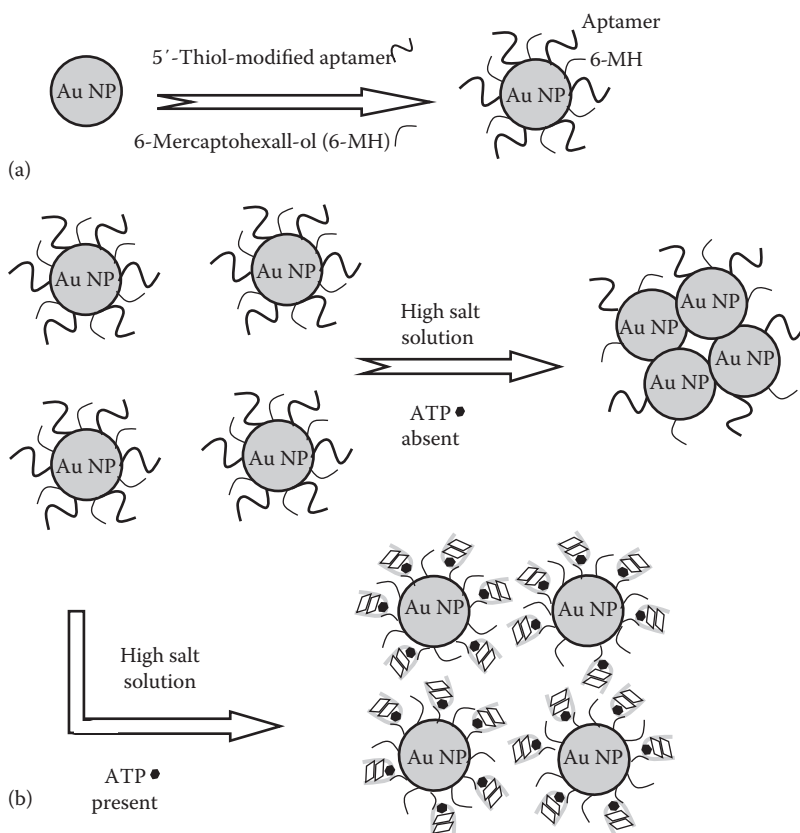
FIGURE 8.20
(a) Synthesis of Aptamer-Au nanoparticles. (b) Sensing mechanism of Apt-Au NPs for the colorimetric determination of ATP: The solution containing no ATP changes color from wine-red to purple due to aggregation while one containing ATP retains its original red-wine color. (After Chen, S.-J. et al., *Biosens. Bioelectron.*, 23, 1749, 2008.)

Chen et al. (2008) described a colorimetric sensing approach for the determination of ATP in urine samples (Figure 8.20). A colorimetric sensor measures the concentration of a known constituent of a solution by comparison with spectroscopic or visual standards. *What is the principle behind this approach?* In the absence of ATP, the color of the Apt-Au NPs solution changed from wine-red to purple via self-induced aggregation. In the presence of ATP, binding of the analytes to the Apt-Au NPs induced folding of the aptamers on the Au NP surfaces into four-stranded quadruplex or tetraplex (fourfold) structures (G-quartet) and/or an increase in charge density. Consequently, the Apt-Au NPs solution was wine-red in color when the analyte was present. G-quartets are atypical nucleic acid structures consisting of a planar arrangement where each guanine is hydrogen bonded by hoogsteen pairing to another guanine in the quartet.

8.8.3 Aptamer-Based Multicolor Fluorescent Gold Nanoprobe for Simultaneous Adenosine, Potassium Ion, and Cocaine Detection

This probe is useful for rapid screening of environmental or security-related molecular targets. Adenosine is a nucleoside, $C_{10}H_{13}N_5O_4$, composed of adenine linked to ribose, that is a structural component of nucleic acids. Cocaine ($C_{17}H_{21}NO_4$) is a mightily addictive central nervous system stimulant that is snorted, injected, or smoked. Zhang et al. (2010) reported a gold nanoprobe for multiplex (two or more signals over a common channel) detection in homogeneous solution, which was suitable for small molecular analytes and functioned in a mix-and-detect fashion. This nanoprobe (Figure 8.21) combined the high specificity of aptamers with the unique fluorescence quenching property of Au NPs; quenching is the suppression of fluorescence by absorption of the stimulating radiation. At the surface of Au NPs, a nanosurface energy-transfer (NSET) effect occurred, which led to long range energy-transfer-based fluorescence quenching.

What is the strategy of their design? In their design, dyes were held in close proximity to the Au NP surface through hybridization of assembled single-stranded DNA (ssDNA) probes and dye-labeled aptamer sequences, causing efficient Au NP quenched fluorescence (OFF state). In the presence of the targets, the aptamer-target binding separated the duplex (having two identical units), liberated the dye-labeled apatmer into solution, and restored the fluorescence.

Three 3'-thiolated DNA strands (P_A, P_K, and P_C, probes for the anti-adenosine aptamer, potassium-specific G-quartet, and anti-cocaine aptamer, respectively), were mixed at equal molar ratio and co-assembled at the surface of Au NPs. The sequences of these three strands were complementary to those of the three aptamers (AA, antiadenosine aptamer; AK, potassium-specific G-quartet; and AC, anti-cocaine aptamer). These three aptamers were labeled with different dyes at the 50 end and then hybridized with their complementary sequences at the surface of Au NPs, which formed the multicolor gold nanoprobe. The choice of three dyes relied on the consideration of their spectral overlap, with anti-adenosine aptamer labeled with R_{ox}, G-quartet with fluoresceinamidite (FAM), and anti-cocaine aptamer with cyanine dye Cy5. In the OFF state, all fluorophores were in close proximity to Au NPs, which resulted in significant fluorescence quenching due to the NSET effect. Fluorescein amidite, abbreviated FAM, is an important dye used in molecular biology, frequently in creation probes to detect the presence of specific nucleic acid sequences. Cyanine dye is any of a class of dyes containing a $-CH=$ group linking two heterocyclic rings containing nitrogen.

Regarding NSET effect, Jennings et al. (2006a and b) reported the first successful application of a dipole surface-type energy transfer from a molecular dipole to a nanometal surface that more than doubled the Förster radius (22 nm) in place of 10 nm for traditional Förster radius value. Also it followed
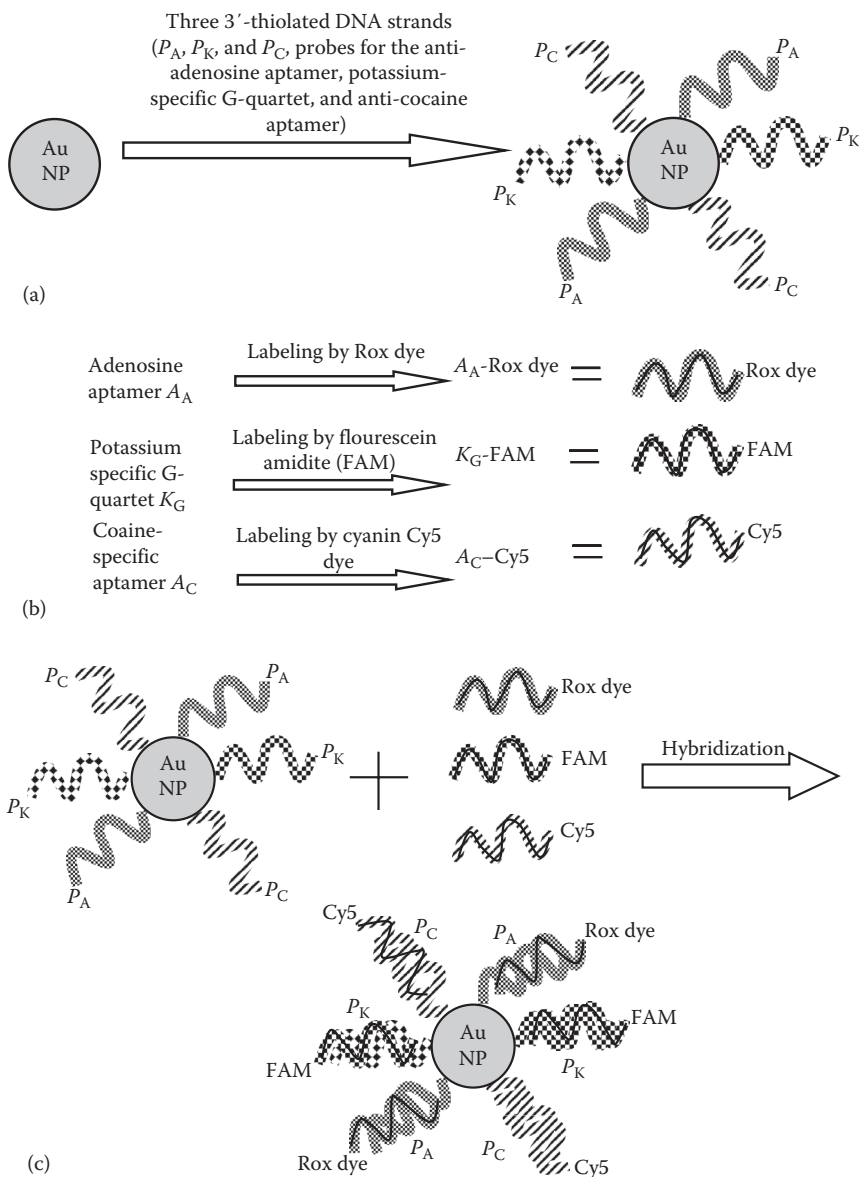
FIGURE 8.21
Multicolor nanosensor for adenosine, potassium and cocaine: (a) Co-assembly of three DNA strands P_A, P_K, and P_C for adenosine, potassium and cocaine on the Au nanoparticle. (b) Labeling of dyes to aptamers: Rox, FAM and Cy5 to adenosine, potassium and cocaine aptamers respectively. (c) Hybridization of dye-attached aptamers with DNA strands affixed to Au nanoparticles. (After Zhang, J. et al., *Small*, 6, 201, 2010.)

$1/R^4$ distance dependence. Because a 1.4 nm gold particle is below the normal mean free-path of an electron in gold at normal temperatures, the only scattering event an electron will feel is that of the surface potential. In other words, it is expected that the electrons spend all of their time at the surface of the particle. If a common fluorophore such as fluorescein ($C_{20}H_{12}O_5$) is placed near such a particle, then the fluorescence quantum efficiency of the dye begins to decrease with a $1/R^4$ distance dependence. The basis of NSET arises from the damping of the fluorophore's oscillating dipole by the free electrons of gold metal as a through-space mechanism. Because the electrons of the Au particle are homogenously oriented, the constraint on dipole–dipole coupling has been greatly relaxed and thus gives rise to energy transfer efficiency at much larger distances.

In the aforementioned experiments, the specific binding of individual aptamers with their specific targets separated the aptamer from the AuNP surface, thus leading to fluorescence recovery that provided quantitative measurement of the analyte concentration. This gold nanoprobe was highly selective since each analyte only displaced the specific aptamer, thus leading to the corresponding dye emission.

8.8.4 Aptamer-Capped QD as Thrombin Nanosensor

For what purpose is thrombin detection required? Thrombosis is an important pathophysiological component of many cardiovascular diseases. Pathophysiology is concerned with the study of the changes of normal mechanical, physical, and biochemical functions, either caused by a disease, or resulting from an abnormal syndrome. Thrombin is a serine protease (an enzyme that cuts certain peptide bonds in other proteins) playing a central role in thrombosis formation (blood clot inside a blood vessel, obstructing the flow of blood through the circulatory system). Detection and imaging of thrombin activity is thus of considerable biomedical interest. Thrombin, also known as coagulation factor II (F2), is a trypsin-like serine protease protein.

Choi et al. (2006) described the application of aptamer-capped quantum dots for thrombin detection (Figure 8.22). 15-mer Thrombin-binding aptamer (TBA) and lead acetate, $Pb(CH_3COO)_2$ were mixed in TAE buffer (pH 8), followed by the injection of sodium sulfide, $Na_2S \cdot 9H_2O$, with vigorous stirring producing TBA-capped QDs which were stable for several months and had diameters of 3–6 nm.

How is thrombin detected? When the aptamer-functionalized QD binds to its target (thrombin), there is a highly selective quenching of the photoluminescence which was ascribed to the charge transfer from functional groups on thrombin to QD. This interaction was highly selective resulting in a NP optical probe for thrombin.
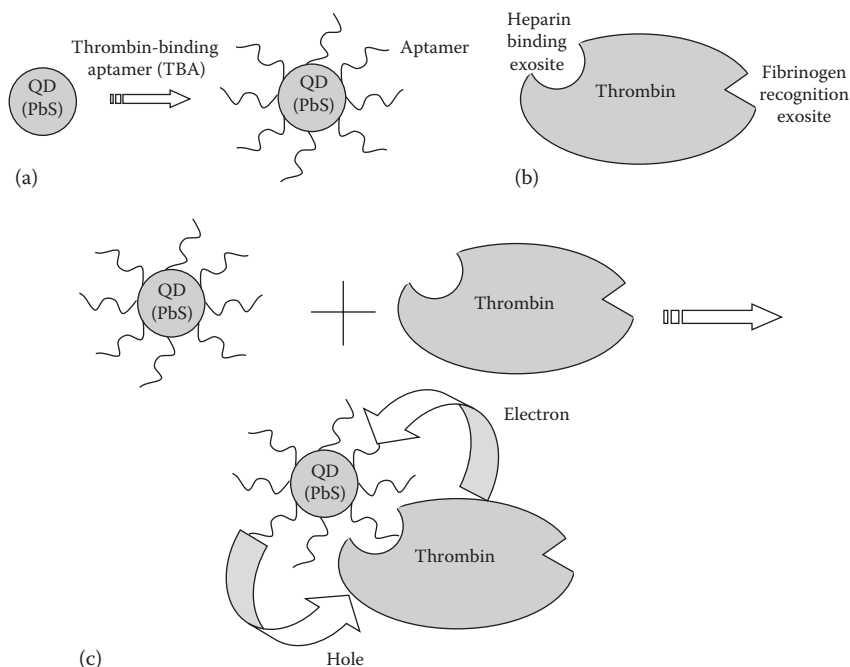
FIGURE 8.22
(a) Functionalization of QD with aptamer forming TBA-capped QD. (b) Blood-clotting human α-thrombin. (c) Interaction of TBA-capped QD with thrombin and induction of photobleaching by charge transfer in the QD–thrombin complex. (After Choi, J. H. et al., *J. Am. Chem. Soc.*, 128, 15584, 2006.)

8.8.5 QD Aptameric Cocaine Nanosensor

Why is cocaine detection necessary? Simple and sensitive detection of cocaine is crucial for law enforcement and clinical diagnostics.

Zhang and Johnson (2009) developed a single-QD aptameric sensor for cocaine (Figure 8.23). The sensor is formed by functionalizing the surface of a QD with an aptamer which can recognize cocaine. 605 QD was used as the energy donor and Cy5 (Cyanine 5: organic compound containing two heterocyclic radicals connected by a chain consisting of an odd number of methine groups) as the acceptor for FRET signal.

What happens in presence of cocaine? The presence of cocaine led to the formation of a cocaine-aptamer complex, which made the Cy 5 labeled oligonucleotide dissociate from the aptamer and 605 QD. The decrease of the Cy 5 signal due to the absence of FRET between 605 QD and Cy 5 signified that cocaine was present. *What are the possible applications of cocaine sensor?* This single QD-based aptameric sensor has potential applications in forensic analysis, environmental monitoring, and clinic diagnostics.
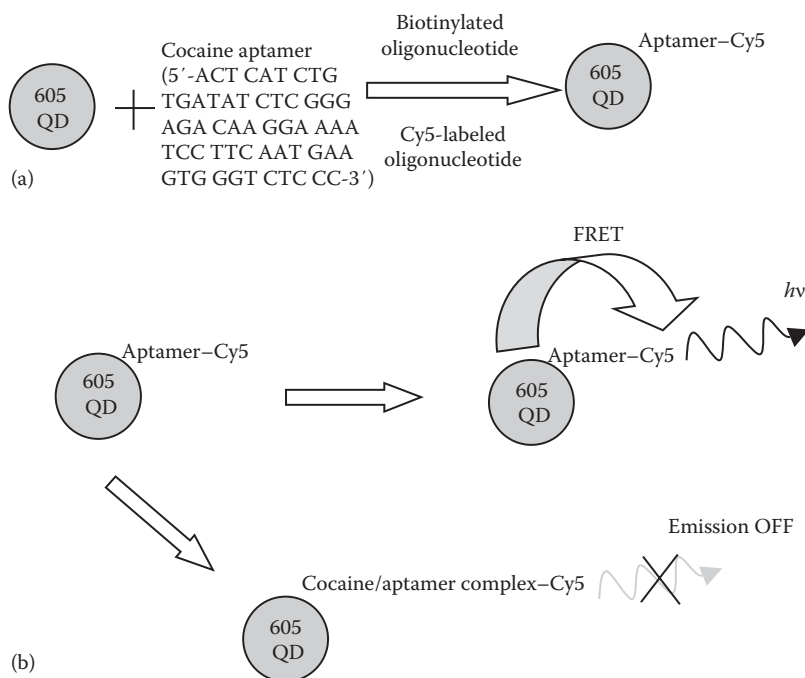
FIGURE 8.23
(a) Functionalization of 605 QD with cocaine-recognizing aptamer. (b) Detection of cocaine with QD-aptameric nanosensor: when cocaine is absent, fluorescence resonance energy transfer (FRET) between 605 QD and Cy5 dye take place accompanied by emission of light, but when cocaine is present, a complex structure is formed between cocaine and the aptamer extinguishing the emission. Abolishment of emission indicates the presence of cocaine. (After Zhang, C.-Y. and Johnson, L.W., *Anal. Chem.*, 81, 3051, 2009.)

8.9 Biochips (or Microarrays)

What is the notion of a biochip? A biochip (biological microchip) can be considered as a biological equivalent for computer microchip or even a microchip that is able to interact with biomolecules (Tarakanov et al. 2010). However, instead of performing millions of mathematical operations, a biochip is intended for carrying out thousands of biological interactions per minute, i.e., thousands times faster in comparison with existing technologies. *How are biochips superior to traditional devices?* (i) The main advantage of biochips over conventional analytical devices is the possibility of massive parallel analysis. (ii) Biochips are smaller than conventional testing systems and highly economical in the use of specimen and reagents.

Some examples of biochips are given in the following: (i) The development of DNA chips was one of the fastest growing areas in DNA/RNA analysis. DNA chips rapidly decrease laboratory turnaround times so that results

are available within 2–6h compared to perhaps 24h. (ii) CNT-FET can be used as a label-free biosensor with a clear change of the CNT conductance in response to the ligand binding event. CNT-based protein biochips are used for the simultaneous detection of several neurotransmitters like DA, catecholamine, serotonin, e.g., a multiwalled-CNT has been synthesized and used for sensitive DA detection (Wu et al. 2007). As repeatedly emphasized, DA and serotonin are important neurotransmitters that interact in the brain. While DA is easily detected with electrochemical sensors, the detection of serotonin is more difficult because reactive species formed after oxidation can absorb to the electrode, reducing sensitivity. The CNT-modified electrodes not only increase the sensitivity and reduce fouling but also allow monitoring DA and serotonin changes simultaneously. These studies show that CNT-coated microelectrodes can be used with fast scanning techniques and are advantageous for *in vivo* measurements of neurotransmitters.

8.10 Discussion and Conclusions

Tables 8.1 through 8.5 compile the vast diversity of biosensor approaches. Electrochemical sensing has received considerable attention. Modified

TABLE 8.1

Biosensors Based on Electrochemical Electrode Modification by Metallic Nanoparticles

Sl. No.	Name of the Nanosensor	Nanomaterial/Nanostructure Used
1.	Nitric oxide (NO) biosensor	(i) Au nanoparticle arrays on nanostructured ITO electrode; (ii) nano-Pt/MWNT modified electrode; (iii) Au nanoparticles-chitosan nanocomposite on a glassy carbon electrode
2.	Dopamine, uric acid and ascorbic acid biosensor	(i) Gold nanoparticle arrays with 3-mercapto-propionic acid (MPA) on ITO; (ii) poly(3-methylthiophene) (P3MT)/gold nanoparticles (AuNPs), on a glassy carbon electrode
3.	CO biosensor	Au nanoparticle cores covered with thiolate monolayers on GC electrodes
4.	Glucose biosensor	GO_x/Au_{coll}–Cyst–AuE
5.	DNA biosensor	Colloidal gold nanoparticles
6.	Allergen–antibody reaction biosensor	Gold nanoparticle–modified graphite carbon electrode
7.	Hepatitis B immunosensor	Pt electrode modified with colloidal gold and polyvinyl butyral
8.	Carcinoembryonic antigen (CEA) biosensor	Mercaptopropanesulfonic-modified gold electrode surface based on $(AuNPs/MWCNTThi/Chit)_n$
9.	*Escherichia coli* biosensor	Screen-printed carbon electrode–modified with ferrocenedicarboxylic acid and AuNPs

TABLE 8.2

Biosensors Based on Electrochemical Electrode Modification by CNTs

Sl. No.	Name of the Nanosensor	Nanomaterial/Nanostructure Used
1.	Dopamine biosensor	(i) Mixing CNTs with bromoform and packing the paste inside a glass tube; (ii) poly (3-methylthiophene) modified glassy carbon electrode (GCE) coated with Nafion/SWCNTs; (iii) fullerene-C60 coated gold electrode
2.	Catalase biosensor	CNT-modified electrode by casting an aliquot of a CNT suspension on the substrate surface
3.	DNA biosensor	(i) MWCNT-modified electrode; (ii) sidewall and end-functionalized CNTs
4.	Glucose biosensor	Nafion overcoated nanotube/GO_x composite-modified electrode
5.	Cholesterol biosensor	Cholesterol esterase, cholesterol oxidase, peroxidase and potassium ferrocyanide immobilized on the screen-printed carbon electrodes and multi-walled carbon nanotubes (MWCNT) added
6.	H_2O_2 biosensor	(i) Immobilization of Hb on multiwall carbon nanotubes and gold colloidal nanoparticles; (ii) entrapping Hb in a composite electrodeposited chitosan-multiwall carbon nanotubes film by assembling gold nanoparticles and hemoglobin step by step

TABLE 8.3

Biosensors Based on QDs

Sl. No.	Name of the Nanosensor	Biomaterial Used
1.	Uric acid biosensor	Uricase-ZnS QD/L-cysteamine
2.	Hydrogen peroxide biosensor	(i) Horseradish peroxidase (HRP) with lipophilic CdSe/ZnS QDs on a glassy carbon electrode; (ii) Hb modified with QDs (CdS) at a normal graphite (GP) electrode
3.	Glucose biosensor	Glucose oxidase (GOD) adsorbed on a CdS nanoparticles–modified pyrolytic graphite electrode
4.	H_2O_2 biosensor	Immobilization of Hb with cadmium sulfide (CdS) nanoparticles (NPs) on pyrolytic graphite (PG) electrode
5.	Light-activated glucose biosensor	Immobilization of CdSe/ZnS nanocrystals on gold electrode

TABLE 8.4

Biosensors Based on Nanowire FETs

Sl. No.	Name of the Nanosensor	Biomaterial/Nanostructure Used
1.	DNA biosensor	ssDNA probe
2.	Protein biosensor	Biotin–SA probe–receptor complex

TABLE 8.5

Biosensors Based on Cantilevers

Sl. No.	Name of the Nanosensor	Biomaterial/Nanostructure Used
1.	DNA biosensor	Two cantilevers with a control (noncomplementary) oligonucleotide in one of them and the DNA probe (complementary) immobilized in the other
2.	Osteosarcoma biosensor	Cantilever sensing the interactions between vimentin antibodies and antigens

electrodes have been prepared by impregnating metallic NPs or CNTs along with biomolecules. Immobilization of biomolecules on QDs has been successfully applied in several cases. Platforms like NW FETs and cantilevers have attracted great interest.

Biosensors are animated with the application of the enlivening biomolecules on suitable platforms. Without these molecules, they are senseless. As the effect of these biomolecules weakens, biosensors lose their senses and ultimately stop functioning. Therefore, biosensors need careful attention to storage and handling. They should also be used with caution bearing in mind their storage environment, their useful shelf lives and other limitations. After all, they rely on living world molecules for their normal functioning.

Review Exercises

8.1 Define the term "biosensor." What are the three main components of a biosensor. Explain with the help of a diagram.

8.2 What type of chemical sensors are classified as biosensors? Justify the statement, "Biosensors constitute a subset of chemical sensors but are studied separately because of their extreme importance."

8.3 What is a nanobiosensor? Bring out the implication of "nano" in this term.

8.4 Why do gold NPs behave as "catalytically active"? Name two methods used to immobilize NPs on electrodes. Name three methods of preparing gold NP-modified electrodes.

8.5 Define the following terms: voltammetry, amperometry, potentiometry, impedometry, and conductometry.

8.6 What is meant by LSV and CV?

8.7 How do gold NPs help in reducing the insulating effect of the protein shell for the direct electron transfer to the electrode?

8.8 Mention some examples where NO sensing is required? Describe the electrochemical sensors for NO detection using nanomaterials.

8.9 What human body abnormalities are detected by determination of DA? How does AA affect DA measurements? Describe the functioning of one gold NP-modified electrode for DA sensing.

8.10 How is direct electron transfer from the enzyme to the electrode inhibited in a glucose biosensor? How do gold NPs help in the immobilization of enzyme? Describe the main steps involved in the preparation of the structure GO_x/Au_{coll}–Cyst–AuE.

8.11 Describe with the help of a diagram the basic principle of a DNA biosensor. How is DNA detected by conductometric method? Illustrate with diagram. Briefly explain electrochemical stripping method of DNA sensing.

8.12 How does the use of polyelectrolyte layers help in discriminating whether the electrochemical signal has originated from silver deposition on background gold electrode surface or the gold NP label? Explain with diagrams the various steps of this scheme for enhanced detection of DNA hybridization.

8.13 How do CNTs function as electrode modifiers in biochemical reactions? Which sites on CNTs are more electroactive? Describe some strategies applied with CNTs for DA measurements.

8.14 How are SWCNTs used for electrical connection between the enzyme redox centers and the electrode? Describe the assembly of the SWCNT electrically contacted glucose oxidase electrode. Highlight the role of edge plane sites on nanotubes in adsorption of glucose oxidase.

8.15 What is the upper limit of blood cholesterol in human beings? Describe a biosensor for monitoring total cholesterol in blood using cholesterol esterase, cholesterol oxidase, peroxidase, potassium ferrocyanide and MWCNTs immobilized on a CPE.

8.16 Name two methods for functionalizing CNTs for biosensors. What are the two types of sites on CNTs for immobilizing biomolecules?

8.17 How Hb modified with quantum dots works as a hydrogen peroxide sensor? What are the useful features of this biosensor? What are the advantages of Hb over HRP?

8.18 How does a QD-based light-triggered glucose sensor operate?

8.19 Bring out the relative merits and demerits of CNTs and NWs for sensor fabrication. Outline the methods for functionalizing Si NWs having a native oxide layer and those with H-terminated silicon surfaces.

8.20 Describe a Si NW-based FET nanosensor for protein detection.

8.21 What constructional materials are used for fabricating microcantilevers? Do uncoated cantilevers exhibit chemical sensitivity? Point out some advantages of using several cantilevers in parallel.

8.22 What criteria are applied for choosing an immobilization process for molecular recognition agents on the cantilever? How are silicon oxide and gold surfaces modified for attaching biomolecules on cantilever surfaces?

8.23 When biotin-containing materials are coated on a cantilever surface and SA is applied to the surface, biotin–streptavidin binding takes place. In one case, the cantilever remained unaffected, in another situation it bent downward while in a still another case, it bent upward on applying SA. Name the three biotin-based compositions that were applied on the cantilever.

8.24 Elaborate the statement: "Gold NPs act as electron transfer wires in many bioelectrochemical reactions."

8.25 How does the immobilization of biomolecules on naked gold surfaces differ from that on gold NPs?

8.26 Mention the advantages of aptamers as compared to antibodies for biosensing applications. Describe the functioning of a QD-based nanosensor for cocaine.

8.27 Describe the use of gold NPs to fabricate an aptamer-based multicolor nanoprobe for selective, multiplex detection of several analytes in homogeneous solution. How does this probe utilize the highly specific binding abilities of aptamers with the ultrahigh quenching ability of Au NPs?

8.28 Explain through schematic diagram the interaction of TBA-Capped PbS QD with thrombin.

8.29 Describe a nanostructured electrode on the basis of CdSe/ZnS quantum dots for the sensitive measurement of the enzyme cofactor NADH. Is glucose detection feasible with such an electrode system and photocurrent measurements?

8.30 Different kinds of NPs, and sometimes the same kind of nanoparticles, can play different roles in various biosensor systems. Elucidate this statement taking the example of "gold nanoparticles."

References

Agarwal, A., V. K. Khanna, and C. Shekhar. 2010. CMOS compatible silicon nanowire arrays for CHEMFET and BioFET (PF-13), *International Conference on Quantum Effects in Solids of Today (I-ConQuEST)*, December 20–23, 2010, NPL, New Delhi, Souvenir & Abstracts, pp. 177–178.

Andreescu, S. and L. A. Luck. 2008. Studies of the binding and signaling of surface-immobilized periplasmic glucose receptors on gold nanoparticles: A glucose biosensor application. *Analytical Biochemistry* 375(2): 282–290.

Balal, K., H. Mohammad, B. Bahareh, B. Ali, H. Maryamc, and Z. Mozhgand. 2009. Zeolite nanoparticle modified carbon paste electrode as a biosensor for simultaneous determination of dopamine and tryptophan. *Journal of the Chinese Chemical Society* 56: 789–796.

Britto, P. J., K. S. V. Santhanam, and P. M. Ajayan. 1996. Carbon nanotube electrode for oxidation of dopamine. *Bioelectrochemistry and Bioenergetics* 41: 121–125.

Butt, H.-J. 1996. A sensitive method to measure changes in the surface stress of solids. *Journal of Colloid and Interface Science* 180: 251–260.

Carrascosa, L. G., M. Moreno, M. Alvarez, and L. M. Lechuga. 2006. Nanomechanical biosensors: A new sensing tool. *Trends in Analytical Chemistry* 25(3): 196–206.

Chen, S.-J., Y.-F. Huang, C.-C. Huang, K.-H. Lee, Z.-H. Lin, and H.-T. Chang. 2008. Colorimetric determination of urinary adenosine using aptamer-modified gold nanoparticles. *Biosensors and Bioelectronics* 23: 1749–1753.

Chen, S., R. Yuan, Y. Chai, L. Zhang, N. Wang, and X. Li. 2007. Amperometric third-generation hydrogen peroxide biosensor based on the immobilization of hemoglobin on multiwall carbon nanotubes and gold colloidal nanoparticles. *Biosensors and Bioelectronics* 22: 1268–1274.

Choi, J. H., K. H. Chen, and M. S. Strano. 2006. Aptamer-capped nanocrystal quantum dots: A new method for label-free protein detection. *Journal of American Chemical Society* 128: 15584–15585.

Cui, Y., Q. Q. Wei, H. K. Park, and C. M. Lieber. 2001. Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species. *Science* 293: 1289–1292.

Curreli, M., R. Zhang, F. N. Ishikawa, H.-K. Chang, R. J. Cote, C. Zhou, and M. E. Thompson. 2008. Real-time, label-free detection of biological entities using nanowire-based FETs. *IEEE Transactions on Nanotechnology* 7(6): 651–667.

Drummond, T. G., M. G. Hill, and J. K. Barton. 2003. Electrochemical DNA sensors. *Nature Biotechnology* 21: 1192–1199.

Fritz, J., M. K. Baller, H. P. Lang, H. Rothuizen, P. Vettiger, E. Meyer, H.-J. Güntherodt, C. Gerber, and J. K. Gimzewski. 2000. Translating biomolecular recognition into nanomechanics. *Science* 288: 316–318.

Geng, D. S. and G. X. Lu. 2007. Size effect of gold nanoparticles on the electrocatalytic oxidation of carbon monoxide in alkaline solution. *Journal of Nanoparticle Research* 9: 1145–1151, doi: 10.1007/s11051-007-9210-1

Goyal, R. N., V. K. Gupta, N. Bachheti, and R. A. Sharma. 2008. Electrochemical sensor for the determination of dopamine in presence of high concentration of ascorbic acid using a Fullerene-C60 coated gold electrode. *Electroanalysis* 20(7): 757–764.

Goyal, R. N., V. K. Gupta, M. Oyama, and N. Bachheti. 2007. Gold nanoparticles modified indium tin oxide electrode for the simultaneous determination of dopamine and serotonin: Application in pharmaceutical formulations and biological fluids. *Talanta* 72: 976–983.

Guiseppi-Elie, A., C. Lei, and R. H. Baughman. 2002. Direct electron transfer of glucose oxidase on carbon nanotubes. *Nanotechnology* 13: 559–564.

Hansen, K. M. and T. Thundat. 2005. Microcantilever biosensors. *Methods* 37: 57–64.

Herne, T. M. and M. J. Tarlov. 1997. Characterization of DNA probes immobilized on gold surfaces. *Journal of American Chemical Society* 119: 8916–8920.

Holloway, A. F., G. G. Wildgoose, R. G. Compton, L. Shao, and M. L. H. Green. 2008. The influence of edge-plane defects and oxygen-containing surface groups on the voltammetry of acid-treated, annealed and "super-annealed" multiwalled carbon nanotubes. *Journal of Solid State Electrochemistry* 12: 1337–1348.

Huang, X., Y. Li, P. Wang, and L. Wang. 2008. Sensitive determination of dopamine and uric acid by the use of a glassy carbon electrode modified with poly(3-methyl-thiophene)/gold nanoparticle composites. *Analytical Sciences* 24: 1563–1568.

Huang, H. Z., Z. G. Liu, and X. R. Yang. 2006. Application of electrochemical impedance spectroscopy for monitoring allergen-antibody reactions using gold nanoparticle–based biomolecular immobilization method. *Analytical Biochemistry* 356: 208–214.

Huang, Y., W. Zhang, H. Xiao, and G. Li. 2005. An electrochemical investigation of glucose oxidase at a CdS nanoparticles modified electrode. *Biosensors and Bioelectronics* 21: 817–821.

Jennings, T. L., J. C. Schlatterer, M. P. Singh, N. L. Greenbaum, and G. F. Strouse. 2006a. NSET molecular beacon analysis of hammerhead RNA substrate binding and catalysis. *Nano Letters* 6(7): 1318–1324.

Jennings, T. L., M. P. Singh, and G. F. Strouse. 2006b. Fluorescent lifetime quenching near $d = 1.5$ nm gold nanoparticles: Probing NSET validity. *Journal of American Chemical Society* 128: 5462–5467.

Ju, H. X., S. Q. Liu, B. X. Ge, F. Lisdat, and F. W. Scheller. 2002. Electrochemistry of cytochrome c immobilized on colloidal gold modified carbon paste electrodes and its electrocatalytic activity. *Electroanalysis* 14: 141–147.

Kerman, K., Y. Morita, Y. Takamura, M. Ozsoz, and E. Tamiya. 2004. DNA-directed attachment of carbon nanotubes for enhanced label-free electrochemical detection of DNA hybridisation. *Electroanalysis* 16(20): 1667.

Khanna, V. K. 2008. New-generation nano-engineered biosensors, enabling nanotechnologies and nanomaterials. *Sensor Review* 28(1): 39–45.

Lee, T. M.-H., L.-L. Li, and I.-M. Hsing. 2003. Enhanced electrochemical detection of DNA hybridization based on electrode-surface modification. *Langmuir* 19: 4338–4343.

Li, G., J. M. Liao, G. Q. Hu, N. Z. Ma, and P. J. Wu. 2005. Study of carbon nanotube modified biosensor for monitoring total cholesterol in blood. *Biosensors and Bioelectronics* 20: 2140–2144.

Li, N. Q., J. X. Wang, and M. X. Li. 2003. Electrochemistry at carbon nanotube electrodes. *Reviews in Analytical Chemistry* 22: 19–33.

Lia, Y., H. J. Schluesenerb, and S. Xu. 2010. Gold nanoparticle-based biosensors. *Gold Bulletin* 43(1): 29–41.

Lin, Y. H., S. Chen, Y. Chuang, Y. Lu, T. Y. Shen, C. A. Chang, and C. S. Lin. 2008. Disposable amperometric immunosensing strips fabricated by Au nanoparticles-modified screen-printed carbon electrodes for the detection of foodborne pathogen *Escherichia coli* O157:H7. *Biosensers and Bioelectronics* 23: 1832–1837.

Liu, C. and J. Hu. 2008. Direct electrochemistry of hemoglobin entrapped in composite electrodeposited chitosan-multiwall carbon nanotubes and nanogold particles membrane and its electrocatalytic application. *Electroanalysis* 20(10): 1067–1072.

Lyons, M. E. G. and G. P. Keeley. 2008. Carbon nanotube based modified electrode biosensors. Part 1. Electrochemical studies of the flavin group redox kinetics at SWCNT/glucose oxidase composite modified electrodes. *International Journal of Electrochemical Science* 3: 819–853.

Maye, M. M., Y. B. Lou, and C. J. Zong. 2000. Core-shell gold nanoparticle assembly as novel electrocatalyst of CO oxidation. *Langmuir* 16: 7520–7523.

Mena, M. L., P. Yáñez-Sedeño, and J. M. Pingarrón. 2005 A comparison of different strategies for the construction of amperometric enzyme biosensors using gold nanoparticle-modified electrodes. *Analytical Biochemistry* 336: 20–27.

Milburn, C., J. Zhou, O. Bravo, C. Kumar, and W. O. Soboyejo. 2005. Sensing inter-
    actions between vimentin antibodies and antigens for early cancer detection.
    Journal of Biomedical Nanotechnology 1(1): 30–38.
Ou, C., R. Yuan, Y. Chai, M. Tang, R. Chai, and X. He. 2007. A novel amperometric
    immunosensor based on layer-by-layer assembly of gold nanoparticles-multi-
    walled carbon nanotubes-thionine multilayer films on polyelectrolyte surface.
    Analytica Chimica Acta 603(2): 205–213.
Ozsoz, M., A. Erdem, K. Kerman, D. Ozkan, B. Tugrul, N. Topcuoglu, H. Ekren, and
    M. Taylan. 2003. Electrochemical genosensor based on colloidal gold nanopar-
    ticles for the detection of Factor V Leiden mutation using disposable pencil
    graphite electrodes. *Analytical Chemistry* 75: 2181–2187.
Park, S. J., T. A. Taton, and C. A. Mirkin. 2002. Array-based electrical detection of
    DNA with nanoparticle probes. *Science* 295: 1503–1506.
Patolsky, F., Y. Weizmann, and I. Willner. 2004. Long-range electrical contacting of
    redox enzymes by SWCNT connectors. *Agnewandte Chemie International Edition*
    43(16): 2113–2117.
Schubert, K., W. Khalid, Z. Yue, W. J. Parak and F. Lisdat. 2009 (Web). Quantum-
    dot-modified electrode in combination with NADH-dependent dehydroge-
    nase reactions for substrate analysis. *Langmuir* 26(2): 1395–1400. doi: 10.1021/
    la902499e.
Shu, W., E. D. Laue, A. A. Seshia. 2007. Investigation of biotin streptavidin binding
    interactions using microcantilever sensors. *Biosensors and Bioelectronics* 22(9–10):
    2003–2009.
Tang, D., R. Yuan, Y. Chai, J. Dai, X. Zhong, and Y. Liu. 2004. A novel immunosensor
    based on immobilization of hepatitis B surface antibody on platinum electrode
    modified colloidal gold and polyvinyl butyral as matrices via electrical imped-
    ance spectroscopy. *Bioelectrochemistry* 65: 15–22.
Tarakanov, A. O., L. B. Goncharova, and Y. A. Tarakanov. 2010. Carbon nanotubes
    towards medicinal biochips. *WIREs Nanomedicine and Nanobiotechnology* 2: 1–10.
Wang, F., X. Deng, W. Wang, and Z. Chen. 2010. Nitric oxide measurement in biologi-
    cal and pharmaceutical samples by an electrochemical sensor. *Journal of Solid
    State Electrochemistry*, doi: 10.1007/s10008-010-1157-y
Wang, J., A. N. Kawde, and M. Musameh. 2003a. Carbon-nanotube-modified glassy
    carbon electrodes for amplified label-free detection of DNA hybridisation.
    Analyst 128(7): 912.
Wang, H. S., T. H. Li, W. L. Jia, and H. Y. Xu. 2006. Highly selective and sensitive deter-
    mination of dopamine using a Nafion/carbon nanotubes coated poly (3-methyl-
    thiophene) modified electrode. *Biosensors and Bioelectronics* 22: 664–669.
Wang, J. X., M. X. Li, Z. J. Shi, N. Q. Li, and Z. N. Gu. 2002a. Direct electrochemistry
    of cytochrome c at a glassy carbon electrode modified with single-wall carbon
    nanotubes. *Analytical Chemistry* 74: 1993–1997.
Wang, J., M. Musameh, and Y. Lin. 2003b. Solubilization of carbon nanotubes by
    Nafion towards the preparation of amperometric biosensors. *Journal of American
    Chemical Society* 125: 2408.
Wang, J., R. Polsky, and D. Xu. 2001. Silver-enhanced colloidal gold electrochemical
    stripping detection of DNA hybridization. *Langmuir* 17: 5739–5741.
Wang, G., J. J. Xu, and H. Y. Chen. 2002b. Interfacing cytochrome c to electrodes with
    a DNA–carbon nanotube composite film. *Electrochemistry Communications* 4:
    506–509.

Wang, Z., Q. Xu, H.-Q. Wang, Q. Yang, J.-H. Yu, and Y.-D. Zhao. 2009. Hydrogen peroxide biosensor based on direct electron transfer of horseradish peroxidase with vapour deposited quantum dots. *Sensors and Actuators B: Chemical* 138(1): 278–282.

Wang, M. K., F. Zhao, Y. Liu, and S. J. Dong. 2005. Direct electrochemistry of micro-peroxidase 11 using carbon nanotube modified electrodes. *Biosensors and Bioelectronics* 21: 159–166.

Wee, K. W., G. Y. Kang, J. Park, J. Y. Kang, D. S. Yoon, J. H. Park, and T. S. Kim. 2005. Novel electrical detection of label-free prostate specific antigen (PSA) using self-sensing piezoresistive microcantilever. *Biosensors and Bioelectronics* 20(10): 1932–1938.

Wu, W., H. Zhu, L. Fan, D. Liu, R. Rennenberg et al. 2007. Sensitive dopamine recognition by boronic acid functionalized multi-walled carbon nanotubes. *Chemical Communications* 23: 2345–2347.

Xiao, Y., F. Patolsky, E. Katz, J. F. Hainfeld, and I. Willner. 2003. Plugging into enzymes: Nanowiring of redox enzymes by a gold nanoparticle. *Science* 299: 1877–1881.

Xu, Y., J. Liang, C. Hu, F. Wang, S. Hu, and Z. He. 2007. A hydrogen peroxide biosensor based on the direct electrochemistry of hemoglobin modified with quantum dots. *Journal of Biological Inorganic Chemistry* 12: 421–427, doi: 10.1007/s00775-006-0198-2

Yanez-Sedeno, P. and J. M. Pingarron. 2005. Gold nanoparticle-based electrochemical biosensors. *Analytical and Bioanalytical Chemistry* 382: 884–886.

Yang, K., H. Wang, K. Zou, and X. Zhang. 2006. Gold nanoparticle modified silicon nanowires as biosensors. *Nanotechnology* 17: S276–S279, doi: 10.1088/0957-4484/17/11/S08

Zhang, L., Z. Fang, G.-C. Zhao, and X.-W. Wei. 2008. Electrodeposited platinum nanoparticles on the multi-walled carbon nanotubes and its electrocatalytic for nitric oxide. *International Journal of Electrochemical Science* 3: 746–754.

Zhang, C.-Y. and L.W. Johnson. 2009. Single quantum-dot-based aptameric nanosensor for cocaine. *Analytical Chemistry* 81: 3051–3055.

Zhang, J., H. P. Lang, F. Huber, A. Bietsch, W. Grange, U. Certa, R. Mckendry, H.-J. Güntherodt, M. Hegner, Ch. Gerber. 2006a. Rapid and label-free nanomechanical detection of biomarker transcripts in human RNA. *Nature Nanotechnology* 1: 214–220.

Zhang, F. F., C. X. Li, X. H. Li, X. L. Wang, Q. Wan, Y. Z. Xian, L. T. Jin and K. Yamamoto. 2006b. ZnS quantum dots derived a reagentless uric acid biosensor. *Talanta* 68: 1353–1358.

Zhang, J. D. and M. Oyama. 2005. Gold nanoparticle arrays directly grown on nanostructured indium tin oxide electrodes: Characterization and electroanalytical application. *Analytica Chimica Acta* 540: 299–306.

Zhang, J. D. and M. Oyama. 2007. Electrocatalytic activity of three-dimensional monolayer of 3-mercaptopropionic acid assembled on gold nanoparticle arrays. *Electrochemistry Communications* 9: 459–464.

Zhang, J., L. Wang, H. Zhang, F. Boey, S. Song, and C. Fan. 2010. Aptamer-based multicolor fluorescent gold nanoprobes for multiplex detection in homogeneous solution. *Small* 6(2): 201–204.

Zhou, H., X. Gan, T. Liu, Q. Yang, and G. Li. 2005. Effect of nano cadmium sulfide on the electron transfer reactivity and peroxidase activity of hemoglobin. *Journal of Biochemical and Biophysical Methods* 64: 38–45.

9

Chemical Nanosensors

9.1 Introduction

How is a chemical nanosensor defined? A chemical nanosensor is an electronic device, consisting of a transducer and a sensitive element whose operation relies on at least one of the physical and chemical properties characteristic of the nanostate (Francia et al. 2009).

What distinguishes a chemical nanosensor from other chemical sensors? Basically, the operation of a chemical nanosensor resembles any other chemical sensor involving charge transference between molecules and a sensitive material, thereby producing an electrical and/or optical signal correlated to the kind and number of molecules. However, contrasting macroscopic sensors, chemical nanosensors derive advantages from the unification of four different features of the nanostate: (i) the quantum confinement effect; (ii) the surface/volume ratio, S/V, typical of a specific surface termination and nanoparticle (NP) doping; (iii) the NP morphology and aggregation; and (iv) the nanomaterial agglomeration state. The term "nanoparticle" here refers, in general, to any kind of structure with at least one of its dimensions in the nanorange. Examples are nanowire (NW), nanodot, and nanotube. These properties improve the sensitive behavior of constructional materials of devices, and high sensitivities have been demonstrated. The field of chemical nanosensors has received tremendous impetus from the synthesis and engineering of materials to create devices that exhibit functionalities specifically originated by their nanostates.

What kinds of materials are commonly used in chemical nanosensors? Due to their interesting structures, nanomaterials afford many opportunities for investigating their sensing behavior by making new types of nanosensing structures. The structure of NWs and nanorods is comparable to each other, which are dominated by a wire-like structure whose diameter varies over a broad range from several tens of nanometers to a micrometer.

What is the difference between NWs and nanorods? The typical length of the NWs ranges from several tens to several hundred micrometers, whereas the nanorods are only several micrometers long.

How do nanotubes differ from NWs? Nanotubes possess wire-like nanostructures but they have hollow cores.

What are the features of nanobelts (NBs)? The belt-like nanostructure has a rectangular cross section. Each NB has a uniform width along its entire

length, and the typical widths of the NBs lie in the range of several tens to several hundred nanometers.

This chapter will provide a review on current research status of chemical sensors based on various new types of nanostructured materials such as nanotubes, nanorods, NBs, and NWs. These nanostructure-based sensors represent authoritative detection platforms for a broad assortment of biological, electrochemical, gas, and pH sensors. *What are the various types of nanosensing devices?* The sensing devices include individual nanostructured sensors, multi-nanostructured sensors, MOSFET-based sensors, nanostructured film sensors, and so forth. These nanosensor devices have a number of key characteristics, including high sensitivity, exquisite selectivity, fast response and recovery, and potential for integration of addressable arrays on a massive scale. These properties set them apart from currently available sensor technologies.

9.2 Gas Sensors Based on Nanomaterials

Why is development of gas sensors currently being carried out intensively? This is because environmental pollution and toxicants in industrial and domestic ambience represent acute problems facing mankind. Gas detection is obligatory in many different fields, e.g., industrial, fuel emission control, automobile exhaust emission control, household security, and environmental pollution monitoring. Gas sensors are utilized in factories, laboratories, hospitals, and almost all technical installations. *What gases have evoked special interest?* Gases of interest include CO_2, CO, NO_2, SO_2, O_2, O_3, H_2, Ar, N_2, NH_3, and H_2O, and organic vapors such as methanol (CH_3OH), ethanol (C_2H_5OH), isopropanol (C_3H_8O), benzene (C_6H_6), and some amines (organic compounds and functional groups that contain a basic nitrogen atom with a lone pair).

What are the gas recognition mechanisms? The mechanisms for recognizing the gases to be determined are the absorption processes, e.g., in metal oxides (MOXs) and carbon nanotubes (CNTs), and specific recognition for the formation of supramolecules (large molecules formed by grouping together or bonding several molecules together) or covalent bonds between the sensor and the analyte, as in some metal complexes. Studies have revealed that sensitivity increases or response time decreases as the film thickness or the particle size of MOXs or organic polymers decreases.

A characteristic feature of gas nanosensors is that they have a transducer system consisting of nanometric materials. On this basis, there are two kinds of devices: (1) In some devices, the sensor and the transducer are different components. (2) In other devices, the sensor also functions as the transducer (as in semiconducting oxides and other systems).

9.3 Metallic Nanoparticle-Based Gas Sensors

While fabricating a gas sensor, why are metal NPs generally spread out on the surface of a substrate? Metal NPs are dispersed on the surface of a substrate to increase the area/volume ratio and favor the adsorption of gases. When they come in contact with the analyte, their electronic properties or the properties of the substrate itself change because the gas molecules adsorb on the metal. The NPs allow greater interaction between the molecules and the substrate, which provides more contact area with the analyte. Generally, the deposition is made by vaporizing the metal precursor (one that precedes and indicates the approach of another), which provides the NPs or films for a subsequent annealing (a heat treatment wherein the microstructure of a material is altered, causing changes in its properties).

Why is hydrogen sensing very important? Hydrogen, being an important source of alternative energy for transportation, residential, and industrial sectors, needs a great deal of attention in research related to its storage, transportation, and handling. One of the major problems is that hydrogen is extremely inflammable and volatile. The presence of only 2%–4% hydrogen in air can cause an explosion. Hence, the research community is groping for a better hydrogen sensor for safety monitoring.

What are the desirable qualities of a hydrogen sensor? For a hydrogen sensor to be efficient and reliable, it should be able to detect hydrogen below its explosive limits. Only then will it be worth using. Generally, group VIII transition metals, Ni, Pd, and Pt nanostructured films, are used for hydrogen detection. *What are the drawbacks of conventional palladium-based sensors for hydrogen?* Macroscopic Pd-based hydrogen sensors suffer from two main shortcomings: (i) Long response times ~0.5 s to several minutes, which is inordinately slow for real-time measurements. (ii) Adsorption of gas molecules, such as CH_4, O_2, and CO on the sensor surfaces causing blockage of the adsorption sites for hydrogen molecules.

How do Pd NW sensors alleviate these problems? Favier et al. (2001) and Walter et al. (2002a,b) investigated Pd NW arrays as hydrogen sensors. Palladium NW hydrogen sensors overcame the above problems. NWs consist of agglomerated Pd grains separated by intergrain nanogaps. On exposure to hydrogen, the gas diffuses into the lattice and reacts with the metal forming a metal hydride (PdH_x). As a result, the wire expands in volume, leading to a partial or total closure of the gaps. Due to reduction or shutting down of gaps between grains, a strong increase in the electrical conductivity is observed. This grain swelling is not completely reversible. On withdrawal of H_2, the original volume of grains is restored but their starting positions are not recovered.

How are Pd NWs fabricated for making a gas sensor? Palladium mesowires and NWs are electrically deposited from aqueous solutions of Pd^{2+} on to the naturally present step edges on highly oriented pyrolytic graphite

(poly-crystalline carbonaceous material that is deposited from the gas phase during pyrolysis of hydrocarbons at temperatures that range from 750°C to 2400°C) surfaces; this process is known as *electrodeposition*. Freshly deposited Pd NWs are separated from the graphite surfaces and transferred on to a glass substrate by means of cyanoacrylate glue, an adhesive substance with an acrylate base that is used in industry and medicine.

9.4 Metal Oxide Gas Sensors

What is the basis of the commercially popular solid-state chemical sensors? Most commercial solid-state chemical sensors are based on appropriately structured and doped MOXs, mainly SnO_2 and ZnO. Additionally, the following oxides show a gas response in their conductivity: Cr_2O_3, Mn_2O_3, Co_3O_4, NiO, CuO, SrO, In_2O_3, WO_3, TiO_2, V_2O_3, Fe_2O_3, GeO_2, Nb_2O_5, MoO_3, Ta_2O_5, La_2O_3, CeO_2, and Nd_2O_3.

What different nanomaterial shapes have been used for building these sensors? Chemical sensors are based on nanotubes, nanorods, NBs, and NWs (Huang and Choi 2007) as follows: (i) In general, nanotube-based sensors include MOX tubes such as Co_3O_4, Fe_2O_3, SnO_2, and TiO_2. (ii) MOX nanorods are ZnO, MoO_3, and tungsten oxide. (iii) As for NB-based sensors, the main attention has been focused on MOXs such as ZnO, SnO_2, and V_2O_5 nanosensors, especially on ZnO-NB sensors. (iv) NWs include In_2O_3, SnO_2, ZnO, β-Ga_2O_3, etc.

How is sensitivity of a gas sensor defined? The definition of sensitivity is different for reducing and oxidizing gases: (i) For reducing gases, sensitivity (S) can be defined as the ratio R_a/R_g. (ii) For oxidizing gases, it is the inverse ratio $S = R_g/R_a$. The symbol R_a stands for the resistance of gas sensors in the reference gas (usually, the air) and R_g stands for the resistance in the reference gas containing target gases.

What physicochemical phenomena take place in these sensors? For several decades, the phenomena participating in gas sensor response have been one of the most interesting subjects in semiconductor gas sensors. The semiconducting nature of some MOXs renders it possible for the electrical conductivity of the material to alter when the gaseous composition of the surrounding atmosphere changes. This happens by the mechanism of charge transfer between surface complexes, such as O^-, O_2^-, H^+, and OH^-, on MOXs with interacting gas molecules. Hence, most of these devices rely on conductivity changes. In principle, they are conductimetric (relating to the measurement of conductivity) nanosensors and are aptly called *chemoresistors*, suggesting resistance variation by chemical means.

The exact fundamental mechanisms that cause a gas-sensing response are still controversial and debatable, but, in effect, they include two phenomena: (i) trapping of electrons at adsorbed molecules and (ii) resultant band

bending induced by these charged molecules are responsible for a change in conductivity.

Why do these gas sensors require an elevated temperature for their functioning? As the operating mechanism for gas sensors, namely, the change in electrical conductivity, requires an activation energy (the least amount of energy needed by a system to initiate a particular process) the classical sensors only operate at high temperatures, in excess of 200°C, thereby increasing the power consumption, an intrinsic and inconvenient requirement tied to these sensors.

What arrangements are commonly made for heating the gas-sensitive material? The sensor (Figure 9.1) consists of a platform made of alumina (Al_2O_3) on the lower surface of which is a spiral platinum heater; it also serves as a temperature sensor. On the top surface lie interdigitated Pt electrodes over which the gas-sensing nanoparticle film is deposited. The ceramic (any of various hard, brittle, heat-resistant and corrosion-resistant materials made by shaping and then firing a nonmetallic mineral, such as clay) thick film hotplate consumes a few watts of power during its operation.

What other versions of hotplates are available? A power-saving hotplate based on low temperature co-fired ceramic (LTCC) technology is shown in Figure 9.2. The CMOS (complementary metal oxide semiconductor)-compatible MEMS hotplate, shown in Figure 9.3, consumes only a few milliwatts of power. It has an embedded polysilicon microheater sandwiched between bottom silicon nitride (Si_3N_4)-thermal silicon dioxide (SiO_2) platform and top plasma enhanced chemical vapor deposition (PECVD) oxide isolation layer. The interdigitated Pt electrodes are patterned over this PECVD oxide layer followed by gas-sensing nanofilm deposition.

Since the adsorption process plays the main role, the physical properties and the shape of the material determine the response of the nanosensor. As already emphasized, nanosized materials have a very favorable area/volume ratio. Therefore, for the same chemical composition, the smaller the nanomaterials are, the more sensitive the sensor is. Higher area/volume ratios favor gas adsorption (and a change in conductivity), decrease the response time, and increase the sensitivity of the device. Because of the aforesaid advantages offered by nanosensors over sensors built with bulk material, NPs and thin films of MOXs (less than 1000 nm thick) have been used to detect a large variety of gases.

It was therefore natural to investigate *whether MOXs in their nanostates, being more chemically reactive than in the bulk form, could provide the required sensitivity at room temperature comparable with bulk oxides at elevated temperatures*. Obviously, this could lower the working temperature and minimize the power requirement. Let us look for some exemplary efforts made in this direction.

Using nano forms of MOXs, several gas sensors have been reported, some of which work at room temperature and others at higher temperatures. But irrespective of what their operating temperature may be, sensitivity of these sensors is enhanced in comparison to those made from bulk MOXs. (i) Law et al. (2002) fabricated individual SnO_2 single-crystal nanoribbons configured as
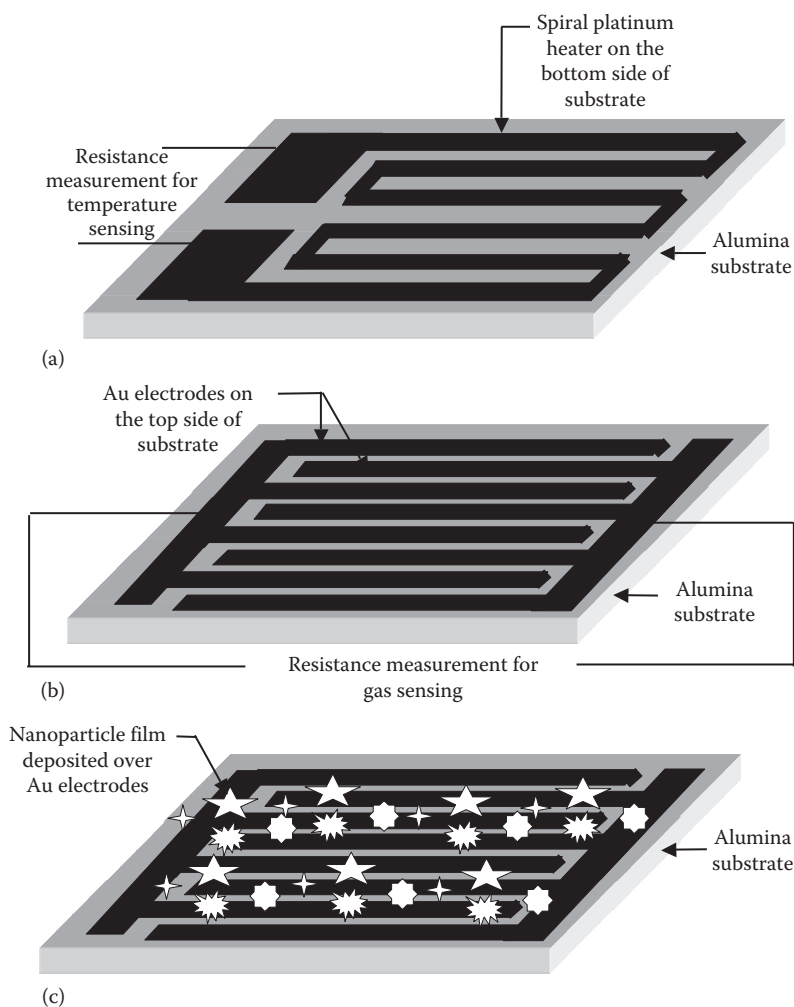
FIGURE 9.1
Gas sensor structure: (a) heater, (b) interdigitated electrodes, and (c) sensing film containing nanoparticles deposited over the interdigitated electrodes.

four-probe conductometric chemical sensors. They reported a detection limit for NO_2 of 3 ppm with response/recovery times ~ seconds. The change in the electrical conductivity was observed even near room temperature. (ii) Fields et al. (2006) developed a hydrogen sensor from single SnO_2 NBs, synthesized via catalyst-free thermal evaporation. The sensitivity and response time of the sensors were measured without any catalyst on the surface to 2% hydrogen at temperatures between 25°C and 80°C. A sensitivity >0.3/(% H_2), with a response time of about 220 s and power consumption of only 10 nW at room temperature, was registered. (iii) Shen et al. (2009) distributed PdO particles randomly on
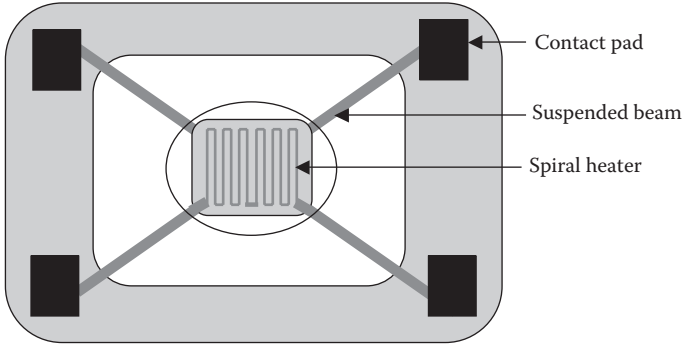
FIGURE 9.2
LTCC hotplate for gas-sensing application. The heater is suspended on a substrate. Heater contact pads are shown as (*H, H*). On the bottom side of substrate lie interdigitated electrodes; these electrodes pass through via holes in the substrate and reach the contact pads (*I, I*).
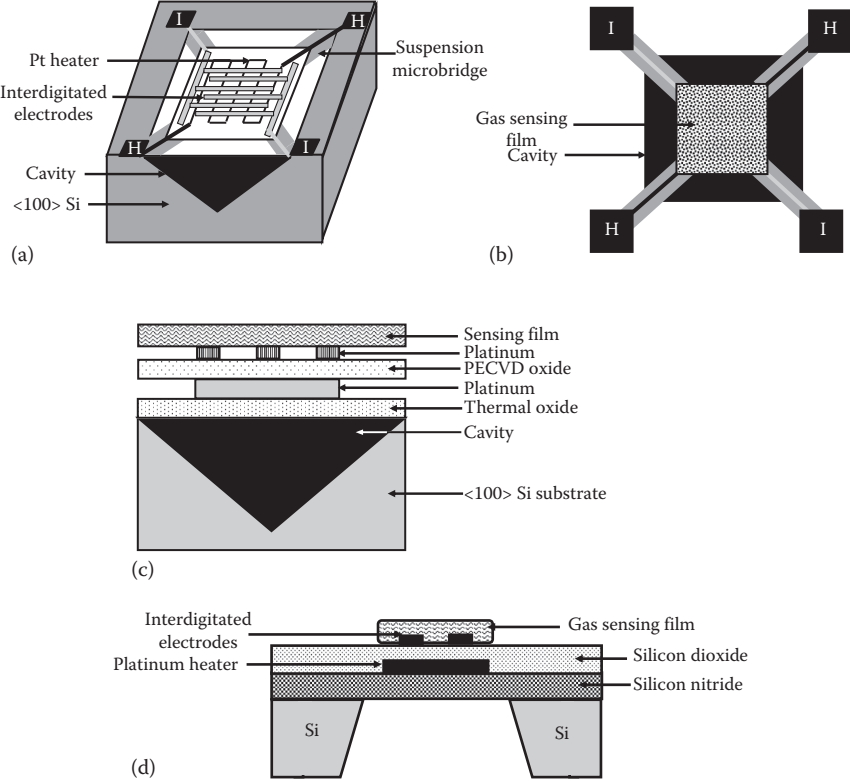
FIGURE 9.3
Schematic representation of (a) MEMS hotplate without gas-sensing film with cavity etched from top side of silicon wafer, (b) its top view with gas sensing film, and (c) its cross-sectional diagram. (d) Cross section of alternative version of MEMS hotplate with cavity etched from bottom side.

the surface of SnO_2 NW testing sensors based on both undoped and Pd-doped SnO_2 NWs. The devices showed a reversible response to H_2 at room temperature, with the signal increasing with rise of Pd concentration.

9.4.1 Sensing Mechanism of Metal Oxide Sensors

These sensors work on the generation of a barrier potential at the MOX surface by gas adsorption? *How is the barrier potential created in a MOX sensor?* As easily understood, response to oxygen is the starting base of semiconductor gas sensors. The sensors are active in air, being vigilant to the target gases. In a first-order elementary analysis, when a MOX crystal such as SnO_2 is heated at a certain high temperature in air, oxygen is adsorbed on the crystal surface acquiring a negative charge, as explained in the following text. Considering oxygen adsorption on a sufficiently large N-type semiconductor crystal, it is known that, if there are traps of electrons on the surface of a semiconductor, conduction electrons are transferred from the subsurface to the traps. Consequently, an electron-depleted layer is left behind. This electron transfer continues until an electronic equilibrium is reached throughout the whole region of the crystal. Thus, donor electrons in the crystal surface are transferred to the adsorbed oxygen, resulting in negative charges on oxygen atoms and leftover positive charges in a space-charge layer. The above-mentioned phenomenon leads to a band bending and the formation of an electron-depleted region (Figure 9.4). Thickness of the electron-depleted region = the length of band bending region. The surface potential thus formed serves as a potential barrier against electron flow (Figure 9.5a), increasing resistivity parallel to the surface. The surface resistance decreases when a reducing gas is present.

How does the sensor resistance decrease in presence of a reducing gas? Inside the sensor, electric current flows through the conjunction parts of SnO_2
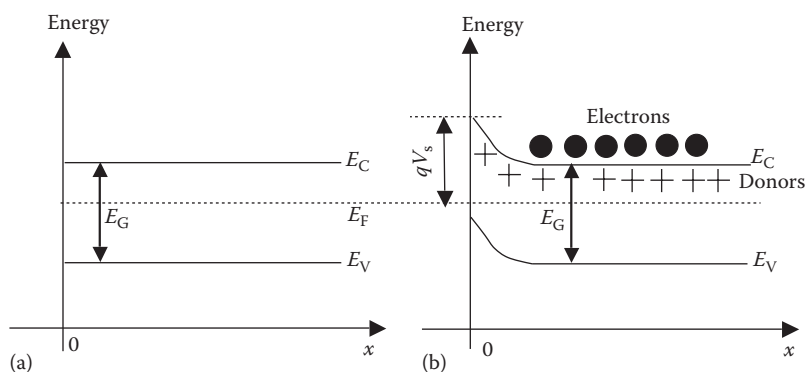
FIGURE 9.4
Energy band diagram, for large N-type semiconductor crystal: (a) before oxygen adsorption and (b) after oxygen adsorption. (After Yamazoe, N. and Shimanoe, K., *J. Electrochem. Soc.*, 155, J85, 2008.)
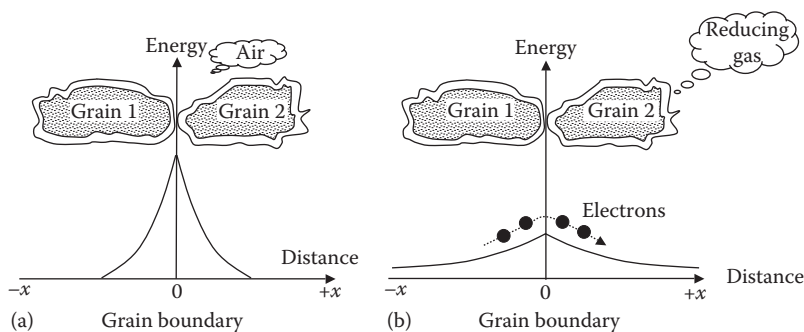
FIGURE 9.5
Intergrain potential barrier: (a) in air and (b) in the presence of reducing gas.

micro/nanocrystals. At grain boundaries, adsorbed oxygen forms a potential barrier, which prevents carriers from moving freely. The electrical resistance of the sensor is attributed to this potential barrier. Reaction of these oxygen species with deoxidizing gases or a competitive adsorption and replacement of the adsorbed oxygen by other molecules decreases the band bending as it tends to curve the bands in the reverse direction. In the presence of a reducing gas, the surface density of the negatively charged oxygen decreases, so the barrier height in the grain boundary becomes smaller (Figure 9.5b). This smaller barrier height decreases overall sensor resistance.

Electrically speaking, what are the two different types of connections between particles of a gas-sensing material? Looking at gas sensor operation more deeply, in sensor devices, constituent crystals are connected to adjacent ones, either by contacts or by necks, the proportion of which depends on the methods and conditions of device fabrication. Spherical crystals (uniform in size), connected with neighbors through a contact or a neck, are believed to be depleted of electrons in the outer region only. There are two different types of contact between the particles of sensing materials: (i) the two-dimensional (2D) contact between necked particles and (ii) point-to-point contact between ordinary particles (Figure 9.6). When the particles are necked together to a large extent and the sizes of the necked part become comparable to the thickness of the resistive electron-depleted layer, the conductive channel through the neck is deterministic of the total resistivity (*neck model*). When the particle sizes are much larger than the thickness of the electron-depleted layer, the conductive channel through the neck becomes too wide. Under these circumstances, the point-to-point contacts between the grain boundaries govern the total resistivity, giving rise to gas sensitivity independent of the particle size.

What are the main models of electron transport between crystals of a gas-sensing material? There are three representative models of electron transport between nearby crystals: (i) When the contact connections prevail, the device
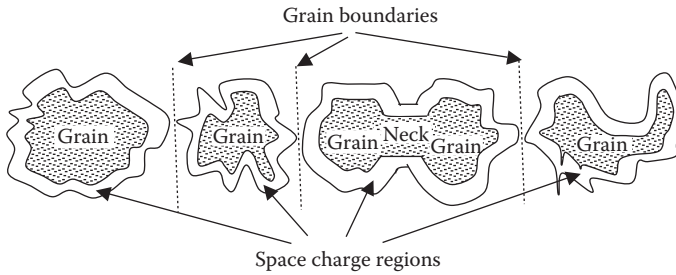
FIGURE 9.6
Neck and grain boundary model of a gas sensor. (After Ma, Y. et al., *J. Wide Bandgap Mater.*, 10, 113, 2002.)

resistance equals the sum of the resistance of each contact. This idea has been conceived in the *double Schottky barrier model*, which presupposes that electrons journey over the barrier at the contact. A Schottky barrier is a potential barrier that is formed in the contact region of a semiconductor, i.e., in the region that adjoins a metal. To obtain a Schottky barrier, the work functions of the metal and the semiconductor must differ. (ii) *Neck* (or *conduction channel model*) allows that electrons travel through a conduction channel that is produced by joining the core regions of bordering spheres. The channel width is tapered at the neck parts so that the conductance is determined by the geometric relation between neck size and depletion depth. (iii) *Tunneling transport model* assumes that electrons located at the periphery of one sphere are transported to that of another by tunneling through a small gap (typically 0–0.1 nm) separating them (Figure 9.7). The conductance is proportional to the density of electrons at the periphery, which is determined by the surface barrier height. In the tunneling model, the height of the wall to tunnel through is set to be equal to the electron affinity of the crystals, i.e., the potential energy difference (V) between the vacuum level and the
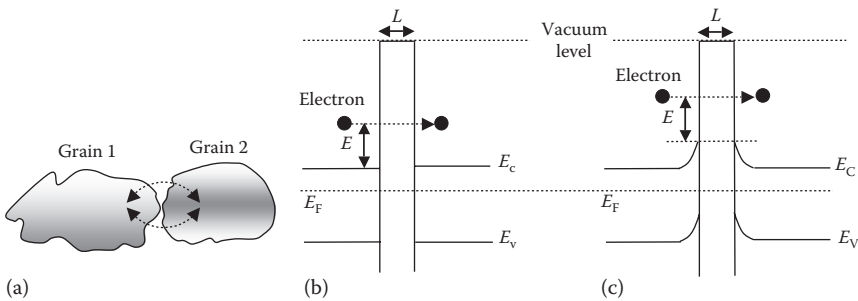
FIGURE 9.7
(a) Electron tunneling conduction across a gap between two grains. Electron tunneling models of gas-sensing material in (b) flat band and (c) bent band conditions. (After Yamazoe, N. and Shimanoe, K., *J. Sens.*, Article ID 875704, 21 pages, doi: 10.1155/2009/875704, 2009.)

conduction band edge at the surface, while the thickness is set to the gap (L) between adjacent crystals.

How does gas adsorption on small crystals differ from that on larger ones? What is meant by regional and volume depletion? Consider the situation after the depletion depth has reached the radius of spheres or necks. Such a situation ensues easily when the spheres are more compact in size or gas adsorption is strengthened. The gas adsorption equilibrium on small crystals is much different from that on large crystals. There are two different cases: (i) In the case of large crystals, depletion does not reach the center of the crystals under usual conditions (*regional depletion*). The intact residual region inside acts as a source of electrons that are supplied to the adsorbates on the surface. In other words, the equilibrium is buffered by the intact region. (ii) In small crystals, depletion can advance beyond just covering the whole region of the crystals to enter into the next stage, *volume depletion*. Equilibration in this stage takes place under the condition that almost all electrons inside the crystals are exhausted. In other words, the equilibrium is close to one in which all electrons available in each crystal are understood to take part evenly. In the stage of volume depletion, the reduced resistance (R/R_0) has a linear relationship with y, the reduced adsorptive strength of the adsorbing gas.

9.4.2 Sensitivity Controlling Parameters and Influence of Heat Treatment

Ma et al. (2002) applied the equation of the semi-infinite planar surface approximation to the grain-boundary potential barrier. Presuming that the resistance in a porous ZnO-based gas element was mainly due to the neck resistance and the grain-boundary resistance, a model of gas sensitivity was developed. This model combining the neck mechanism and the grain boundary mechanism illustrated the effects of neck and grain boundaries on sensitivity. The grain-boundary resistance was derived as

$$R_{GB} = \left\{ \frac{\sqrt{2\pi\alpha mkT}}{A_{GB}q^2 n_b} \right\} \exp\left(\frac{qV_s}{kT} \right) \tag{9.1}$$

where
  α is the ratio of effective electron mass to electron mass
  m is the mass of an electron
  k is the Boltzmann constant
  T is the absolute temperature
  A_{GB} is the cross section of the grain boundary
  q is the electronic charge

n_b denotes the free-electron density in the grain body

qV_s is the height of the grain-boundary potential barrier given by

$$qV_s = \frac{q^2 n_t^2}{2\varepsilon_0 \varepsilon_r n_b} \tag{9.2}$$

where

ε_0 is the free-space permittivity

ε_r is the permittivity of ZnO

n_t is the surface electron density

$$n_t = n_0 - n_r \tag{9.3}$$

where

n_0 is the density of the electrons trapped on the surface by the initially adsorbed oxygen

n_r is the chemisorbed gas concentration

Example 9.1

Given the following parameter values: $\alpha = 0.018$, $m = 9.11 \times 10^{-31}$ kg, $k = 1.3806503 \times 10^{-23}$ m^2 kg s^{-2} K^{-1}, $T = 570$ K, $q = 1.6 \times 10^{-19}$ C, grain size $L = 30$ nm, $n_b = 2.1 \times 10^{14}$ m^{-3}, $n_0 = 1.26 \times 10^{16}$ m^{-2}, $\varepsilon_0 = 8.854 \times 10^{-12}$, $\varepsilon_r = 8$, find the potential barrier and grain-boundary resistances at two values of chemisorbed gas concentration: 1×10^{15} m^{-2} and 1×10^{16} m^{-2}.

From Equation 9.2,

$$V_s = \frac{q n_t^2}{2\varepsilon_0 \varepsilon_r n_b} = \frac{1.6 \times 10^{-19} \times (n_0 - n_r)^2}{2 \times 8.854 \times 10^{-12} \times 8 \times 2.1 \times 10^{24}}$$

$$= 5.378 \times 10^{-34} \times \left(1.26 \times 10^{16} - n_r\right)^2 \tag{9.4}$$

At $n_r = 1 \times 10^{15}$ m^{-2},

$$V_s = 5.378 \times 10^{-34} \times \left(1.26 \times 10^{16} - 1 \times 10^{15}\right)^2 = 5.378 \times 10^{-34}$$

$$\times \left(1.16 \times 10^{16}\right)^2 = 0.07237 \text{ J} \tag{9.5}$$

Now, $\alpha = 0.018$, $m = 9.11 \times 10^{-31}$ kg, $k = 1.3806503 \times 10^{-23}$ m^2 kg s^{-2} K^{-1}, $T = 570$ K, Cross-sectional area of grain boundary is $A_{GB} = \pi(L/2)^2 = 3.14 \times (30 \times 10^{-9}/2)^2 = 7.065 \times 10^{-16}$ m^2, $q = 1.6 \times 10^{-19}$ C, $n_b = 2.1 \times 10^{24}$ m^{-3},

$$R_{GB} = \left\{ \frac{\sqrt{2\pi\alpha m k T}}{A_{GB}q^2 n_b} \right\} \exp\left(\frac{qV_s}{kT} \right)$$

$$= \left\{ \frac{\sqrt{2 \times 3.14 \times 0.018 \times 9.11 \times 10^{-31} \times 1.3806503 \times 10^{-23} \times 570}}{7.065 \times 10^{-16} \times \left(1.6 \times 10^{-19}\right)^2 \times 2.1 \times 10^{24}} \right\}$$

$$\times \exp\left(\frac{1.6 \times 10^{-19} \times 0.07237}{1.3806503 \times 10^{-23} \times 570} \right)$$

$$= \frac{2.8468 \times 10^{-26}}{3.7981 \times 10^{-29}} \times \exp(1.47136) = 7.4953 \times 10^2 \times 4.355 = 3.264\ k\Omega \quad (9.6)$$

At $n_r = 1 \times 10^{16}\,\text{m}^{-2}$,

$$V_s = 5.378 \times 10^{-34} \times \left(1.26 \times 10^{16} - 1 \times 10^{16}\right)^2 = 5.378 \times 10^{-34}$$

$$\times \left(2.6 \times 10^{15}\right)^2 = 6.636 \times 10^{-3}\ J \quad (9.7)$$

$$R_{GB} = \frac{2.8468 \times 10^{-26}}{3.7981 \times 10^{-29}} \times \exp\left(\frac{1.6 \times 10^{-19} \times 6.636 \times 10^{-3}}{1.3806503 \times 10^{-23} \times 570} \right)$$

$$= 7.4953 \times 10^2 \times \exp(0.1349) = 7.4953 \times 10^2\ \Omega \times 1.1444 = 0.8578\ k\Omega \quad (9.8)$$

If the neck is a cylinder of radius r_m, the neck resistance is

$$R_N = \frac{L}{\pi q \left\{ \mu_b n_b r_0^2 + \mu_d n_d \left(r_m^2 - r_0^2 \right) \right\}} \quad (9.9)$$

where
   L is the grain size
   μ_b is the electron mobility in the neutral grain body
   μ_d is the electron mobility in the depleted layer
   n_d is the free-electron density in the depleted layer

$$r_0 = r_m \left\{ 1 - \frac{2n_t}{n_b r_m} \right\}^{0.5} \quad (9.10)$$

r_m is the radius of the cylinder constituting the neck.

    The gas sensitivity will increase with decreasing grain size. Gas sensors fabricated with nanometer ZnO have high sensitivity when the width of the space-charge layer of the neck in air is comparable with

the neck radius. Although the grain-boundary resistance may be $\ll$ than the neck resistance, it cannot be neglected. It was suggested that the decreasing of the ratio between the numbers of grain boundaries and necks was a possible approach to the development of nano-ZnO gas sensors with a high sensitivity.

What factors control the sensitivity of a gas sensor? In accordance with Yamazoe (2005), sensing performances, especially sensitivity, are controlled by three independent factors. These factors are receptor function, transducer function, and utility (Figure 9.8). (i) *Receptor function* concerns the ability of the oxide surfaces to interact with the target gas. Chemical properties of the surface oxygen of the oxide itself are responsible for this function in a clean oxide device, but this function can be, for the most part, modified to induce a large change in sensitivity when an additive (noble metals, acidic, or basic oxides) is loaded on the oxide surface. (ii) *Transducer function* concerns the ability to convert the signal caused by chemical interaction of the oxide surface (work function change; work function is the minimum amount of energy required to remove an electron from the surface of a metal) into electrical signal. This function is executed by each boundary between grains, to which a double-Schottky barrier model is applied. The resistance depends
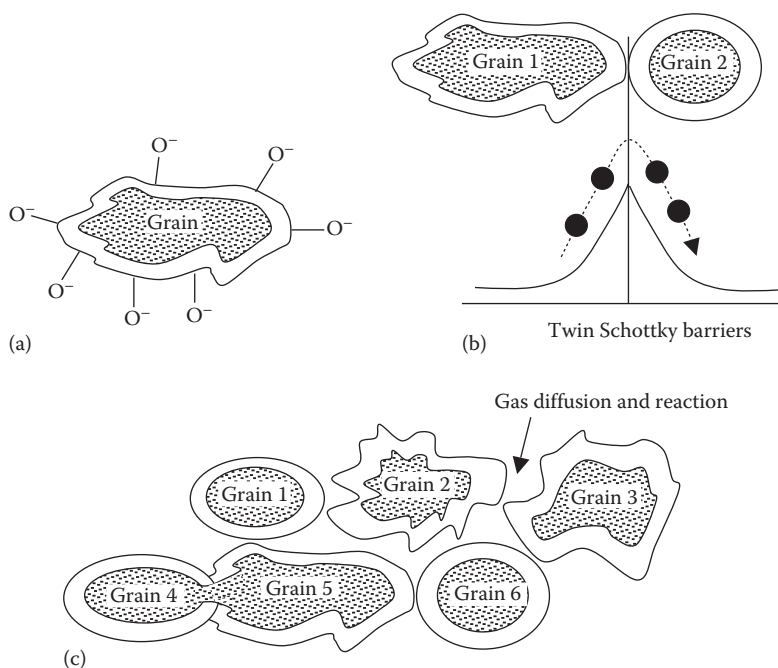
FIGURE 9.8
Three sensitivity-controlling parameters of a gas sensor after Yamazoe (2005): (a) receptor, (b) transducer functions, and (c) utility factor.

on the barrier height and then on the concentration of the target gas. This situation is essentially unchanged with a change in the grain size (diameter, D) of the oxide unless D is kept above a critical value (D_c), which is just equal to twice the thickness (L_s) of surface space-charge layer of the oxide. For values of D smaller than D_c (6 nm for SnO_2), sensitivity increases sharply with decreasing D.

The last factor, *utility*, concerns the accessibility of inner oxide grains to the target gas. The importance of this factor is made obvious when one considers that the target gas (reducing gas) reacts with the oxide surface on the way of diffusing into the bulk of device. If the rate of reaction is too large compared with that of diffusion, the gas molecules cannot access the grains located at inner sites, leaving them unutilized for gas sensing and thus resulting in a loss in sensor response. The existence of this factor was conjectured long time ago from familiar volcano-shaped correlations between sensor response and operating temperature, but quantitative understanding of it was made possible only recently for thin-film devices derived from SnO_2 sols.

What is the role of diffusion rates of gas through the material in the sensitivity of a gas sensor? The sensitivity of a gas sensor is intimately related to the degree of diffusion of gas molecules into the sensing material used to adsorb and desorb them, as well as to the specific surface areas of the latter. *How are these two factors taken into account for sensitivity maximization?* In order to maximize its sensitivity, many endeavors have been made lately to design the structure of the resistive channels so as to enhance the gas diffusion. Although channels with 2D film structures have been widely investigated, their small specific surface areas limit their adsorption efficiency. Recently, channels with one-dimensional (1D) structures, such as NWs, nanorods, and NBs, were proposed. These channels offer a greater degree of diffusion for gas molecules, as well as a relatively larger specific surface area, compared with 2D structured films. These 1D structures provide well-defined channels without any grain boundaries, which might barricade the flow of electrons through the channels. Due to these favorable properties, gas sensors with channels made of 1D nanomaterials have been widely investigated. In addition, sensing materials with zero-dimensional (0D) structures such as NPs have also been researched for use in the channels, since their larger specific surface areas compared to those of the other dimensional structures will hopefully dramatically enhance the sensitivity characteristics of the gas sensors. However, the relatively high temperature (600°C) of the heat treatment used to stabilize the NPs brings about their agglomeration. This causes two critical problems for the gas sensors: (i) The sensitivity decreases, since the gas-sensing reaction occurs mainly at the surface of the agglomerates and (ii) the response is sluggish since the diffusion of gas molecules into the agglomerates becomes more difficult and slower.

Jun et al. (2009) proposed gas sensors with channels composed of non-agglomerated, necked NPs to maximize their sensitivity and response

characteristics. *What are the main ideas of this approach?* Non-agglomerated, necked NPs have the characteristics of a mixture of 0- and 1D structures. If they are slightly necked with the adjacent NPs without any agglomeration, all of them participate in the gas-sensing reaction and, thus, the gas sensitivity is enhanced. In addition, the rapid diffusion of the target gas and speedy counter diffusion of the product gas are enabled and, thus, the response characteristics are improved. The slight necking of the NPs with their neighbors also enhances the conductivity of the channels, due to the lowering of the potential barriers present in the interfaces of the adjacent NPs. Hence, channels composed of non-agglomerated, necked NPs have the merits of both 0D structures with the largest possible specific surface areas and 1D structures with well-defined channels.

How were the ideas implemented and their benefits verified? The above workers pasted the NPs across the electrodes and subsequently heated at 400°C. The sensing characteristics of the gas sensors with channels composed of the heat-treated NPs as sensing materials were then investigated and compared with those of gas sensors made with channels composed of the as-pasted NPs. When exposed to 5 ppm NO_2 gas, the response of the gas sensor with the channels composed of the heat-treated ZnO NPs was about 400, while that of the gas sensors with the channels composed of the as-pasted ZnO NPs was about 150. The higher response of the gas sensors composed of the heat-treated ZnO NPs results from the necking of the neighboring NPs and the size distribution of the NPs. The response and recovery characteristics of the gas sensors developed were remarkably fast at an NO_2 concentration of 0.4 ppm. Their response and recovery times were 13 and 10 s, respectively. Therefore, the low-temperature heat treatment of NPs is one of the useful methods of fabricating gas sensors having a high and fast response. These observations indicate that the non-agglomerated necking of the NPs induced by the heat treatment remarkably boosts the gas-sensing performance of the NP-based gas sensors.

9.4.3 Effect of Additives on Sensor Response

What is the effect of additives on sensor response? The ability of noble metals to act as highly effective oxidation catalysts can be used to enhance the reactions on gas sensor surfaces. A wide diversity of methods, including impregnation, sol–gel, sputtering and thermal evaporation, has been used for introducing noble metal additives into oxide semiconductors. The sol–gel process generally involves the use of metal alkoxides, which undergo hydrolysis and condensation polymerization reactions to give gels. The sol–gel process comprises solution, gelation, drying, and densification. The preparation of a silica glass begins with an appropriate alkoxide, which is mixed with water and a mutual solvent to form a solution. Hydrolysis leads to the formation of silanol groups (Si–OH). These species

are only intermediates. Subsequent condensation reactions produce siloxane bonds (Si–O–Si). The silica gel formed by this process leads to a rigid, interconnected 3D network consisting of submicrometer pores and polymeric chains.

Thus sol–gel is a colloidal suspension of silica particles that is gelled to form a solid. The resulting porous gel is chemically purified and consolidated at high temperatures into high purity silica. Removing the liquid located within the pores leads to a dried gel named xerogel.

Referring back to the discussion on the effect of additives, usually L_s is a function of the concentration of electron donors in the bulk oxide, and D_c is changed by doping the base oxide with a foreign oxide. When the oxide is loaded with a foreign additive, the additive can modify L_s also if it interacts electronically with the oxide. Indeed, such a change in L_s or barrier height elucidates the marked sensitizing effects of certain noble metals like Pd for the sensors of this type, e.g., in the case of Pd-loaded SnO_2, upon exposure to air, Pd is oxidized into PdO, which acts as a strong acceptor of electrons from SnO_2. In this state, each grain of SnO_2 is covered with a strongly electron-deficient space-charge layer, leading to a high resistance. Whenever there is contact with a combustible gas in air, PdO is reduced to Pd, which is no longer an electron acceptor, resulting in a sharp plummeting of the electrical resistance.

What are the mechanisms producing this behavior? It is noted that the sensitizing effects emerge through coupling a redox modification of the additive with a change in its electronic interaction with the oxide grains. Two concepts are hypothesized to explain the improvement of the sensing performance of NW upon Pd deposition: (i) the "electronic mechanism" proposal and (ii) a "chemical process." The *electronic mechanism* considers that depletion zones are formed around the modified particles. Modulation of the nano-Schottky barriers (and hence the width of the conduction channel) occurs. This is due to changes in the oxidation state of the Pd accompanying oxygen adsorption and desorption. The modulation is responsible for the sensing enhancement. Explanation of the kinetics and temperature dependence brought about by Pd functionalization is difficult through the electronic mechanism, while the chemical process does not experience such problems. The mechanism of chemical process is based on the highly effective dissociation catalytic ability of Pd. The ionosorption of oxygen takes place at defect sites of the pristine surface. Pd is a better oxygen dissociation catalyst than tin oxide, catalytically activating the dissociation of molecular oxygen. Then atomic products diffuse to the MOX support. Furthermore, it is not necessary for molecular oxygen to dissociate on Pd surface only. It is believed that oxygen molecules reside briefly on an oxide support and diffuse to a catalyst particle before they have any occasion for desorption. This is called "the back-spillover effect." It increases the sensitivity along with diffusion of atomic products to the MOX support.

9.5 Carbon Nanotube Gas Sensors

9.5.1 Gas-Sensing Properties of CNTs

What properties of CNTs are useful for gas sensing? CNTs are very interesting in the sensor field because they are characterized by high, theoretically infinite, surface/volume ratio (Zhang and Zhang 2009). The development of CNT-based gas sensors has attracted all-embracing interest because of their high response, sub-ppb concentration detection levels, prompt response, low operating temperature, small power consumption, miniature size for nanodevice shrinking, functionalization for specific detection of molecules, and ability to build massive sensor arrays. The development of CNT nanosized sensors opens application possibilities of multi-transducer and multisensor array by using pattern recognition techniques for efficient chemical gas analysis.

What are the chief action lines for realizing CNT gas sensors? There are two approaches to CNT gas sensor fabrication: (i) resistive gas sensor and (ii) field-effect gas sensor. Figure 9.9a shows the adsorption of gas molecules on a CNT surface producing a conductance change. A resistive CNT gas sensor is shown in Figure 9.9b. It is fabricated on a silicon substrate on which an insulating thermal oxide layer is grown. Bundles of CNTs are cast over the two electrodes. The resistance change of CNT bundles with ambient gas concentration is the sensing parameter. A field-effect CNT gas sensor (Figure 9.10) consists of a single semiconducting CNT spanning over the source and drain electrodes. It serves as the channel of the field-effect transistor (FET) device. CNT conductivity varies with gas concentration. A silicon back gate is used to produce the field-effect action. The field-effect gas sensor will be discussed in more detail later.

Several CNT-based nanosensors for gases have been investigated because interaction of molecules adsorbed on their surfaces affects their electrical properties (basically their conductivity). Studies have been made of how they
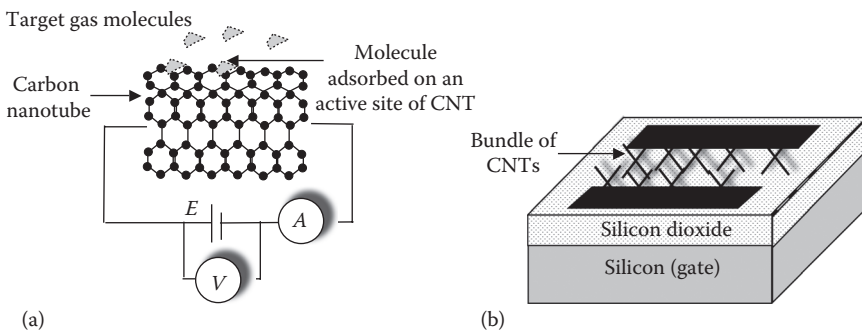
FIGURE 9.9
(a) Basic mechanism of a resistive CNT gas-sensing device and (b) a resistive gas sensor using CNT bundles or thin film as sensing element.
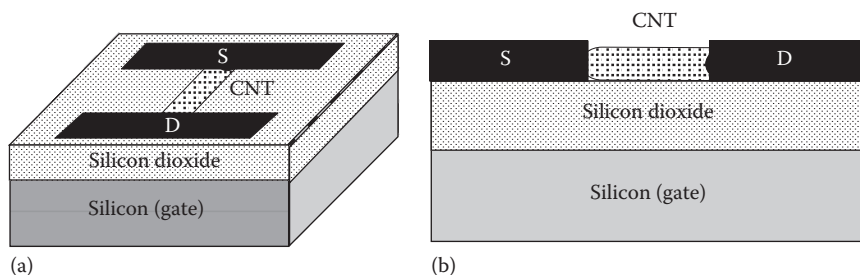
FIGURE 9.10
(a) 3D view of a gas sensor based on FET with a single carbon nanotube as sensing element and (b) its cross-sectional diagram.

behave when exposed to gases such as NO_2, NH_3, CO, H_2O, N_2, or organic vapors. Nanotubes respond differently to different gases that add or remove electron density from the surface of the nanotubes. Single-walled carbon nanotubes (SWCNTs) are sensitive to NO_2, NH_3 and volatile organic compounds. The adsorption of gaseous molecules either donates or withdraws electrons from the SWCNTs, leading to changes of electrical properties of the SWCNTs. The drawback of these sensors is slow and incomplete recovery.

Li et al. (2003) proposed a simple resistive device fabricated by casting a solution of purified SWCNT in dimethyl formamide, $(CH_3)_2NC(O)H$, on a silicon substrate. The sensor worked at room temperature and offered a very high sensitivity to NO_2 and nitrotoluene $(C_7H_7NO_2)$ with 44 and 262 ppb detection limits, respectively, in N_2.

To detect diisopropyl methylphosphonate (DIMP), $C_7H_{17}O_3P$, and dimethyl methylphosphonate (DMMP), $C_3H_9O_3P$, simulants for the G series nerve agents organophosphorus compounds Soman $(C_7H_{16}FO_2P)$ and Sarin, $(C_4H_{10}FO_2P)$, respectively, Cattanach et al. (2006) described the use of network films of SWCNT bundles on flexible substrates such as polyethylene terephthalate (PET), $(C_{10}H_8O_4)_n$. Upon exposure to DIMP or DMMP vapors, large reproducible resistance changes (75%–150%), were observed. Concentrations as low as 25 ppm could be measured. These SWCNT/PET films were readily synthesized using line patterning, which is a simple and rapid room temperature method (involving no photolithography or printing), to obtain any desired pattern of SWCNT films from aqueous surfactant-supported dispersions of SWCNTs on flexible substrates like plastic, paper, cloth, etc.; a surfactant is a surface active agent that lowers the surface tension of a liquid, increasing the contact between the liquid and another substance. The sensor functioned robustly even when the SWCNT/PET sensor was bent all the way to a crease.

Network films of SWCNT/PET are also able to sense DIMP and DMMP when other interferent vapors are present. In addition to air, humidity, etc., network films of SWCNT/PET are able to detect simulant vapors in the presence of large quantities of interferent vapors commonly found in

field applications, such as hexanes (a colorless flammable liquid hydrocarbon with the chemical formula C_6H_{14}) and xylenes, C_8H_{10}, $C_6H_4(CH_3)_2$, or $C_6H_4C_2H_6$, constituents of gasoline (a volatile mixture of flammable liquid hydrocarbons derived chiefly from crude petroleum), and diesel (a mix of hydrocarbons typically in the C_9H_20 to $C_{12}H_{26}$ range derived from petroleum but heavier than gasoline/petrol) fuel, respectively. In order to improve the selectivity of SWCNT/PET films to simulants in a mixture containing large amounts of interferent vapors, the researchers screened a wide range of polymer coatings for chemoselectivity (the preferential reaction of a chemical reagent with one functional group in the presence of other similar functional groups). It was found that a 2 µm thick film of polyisobutylene (PIB), $-(CH_2-C_3H_6)_n-$, spin-coated) on the SWCNT sensor was an effective barrier coating. The PIB-coated sensor entirely screened out 30,000 ppm of interferent vapors. At the same time, it was able to detect 299 ppm DIMP. This showed that the use of PIB as a barrier coating polymer could screen out high concentrations of interferent vapor signatures.

9.5.2 Responses of SWCNTs and MWCNTs

Do SWCNTs, multi-walled carbon nanotubes (MWCNTs), and aligned CNTs differ in their gas-sensing behaviors? Sensors have been fabricated using single-walled, multi-walled, and well-aligned CNT arrays (Penza et al. 2004). Suehiro et al. (2005) demonstrated that the normalized response of the SWCNT sensor was higher than that of the MWCNT sensor. Probably it is because SWCNTs contained more semiconducting tubes. In general, MWCNTs show a conducting (metallic) behavior at room temperature. However, MWCNTs could contain some semiconducting tubes among predominantly metallic ones. The higher normalized response of the SWCNT sensor may be attributed to higher abundance of the semiconducting tube, which is responsible for the sensor response. *Does the alignment of CNTs affect the sensor response?* Yes, nanotube alignment also affects the sensor response, e.g., the response to ammonia (NH_3) of a sensor with aligned SWCNT was double with respect to disordered SWCNT (Lucci et al. 2005). This effect was probably induced by the fact that the ordered SWCNTs were more uniformly exposed to the NH_3 molecules, than in the case of placement of SWCNTs in form of a random network.

9.5.3 Modification of CNTs

Why CNTs need to be modified for making sensors for particular gases? Sensitivity of CNTs to the surrounding chemical environment is advantageous but on the opposite side, it implies a distinct drawback because it means a lack of selectivity. Sensors based on raw CNTs are devices that operate at ambient temperature and are highly sensitive to different types of gases, although they are not very selective. For improving the selectivity of CNT-based sensors, nanotube functionalization has been done.

For this purpose, several functionalization methods have been developed. These include doping or inserting atoms or functional groups into nanotube structures, which constitutes a type of covalent functionalization. Functionalization can also be noncovalent. This consists of depositing particles or coating the surface of the CNT with a material that interacts with the analyte of interest. In the latter case, the nanotubes still act as a transducer system, while the material deposited on the surface becomes the sensing part of the device. The sensitivity of some sensors has been dramatically enhanced from ppm to ppb level after surface modification or functionalization and doping.

What attempts have been made toward CNT functionalization? CNTs have been modified by many workers using functional groups, metals, oxides, polymers, or doping CNTs with other elements to enhance the response and selectivity of the sensors:

1. Lu et al. (2004) reported a SWCNT-based gas sensor, designed for methane (CH_4) with an operating range in dry air between 6 and 100 ppm, fabricated by casting, using a suspension of purified SWCNT in deionized water (DIW) where sputtered Pd NPs were dispersed.

2. Li et al. (2006) fabricated a chemiresistor with an interdigitated electrode (IDE) configuration. Purified SWCNTs were laid across the electrodes using a solution casting process in the form of a network. Polymer coating of SWCNTs allowed selective sensing of certain gases as demonstrated here for chlorine (Cl_2) and hydrochloric acid (HCl) vapor. These are industrially important chemicals, and sensors for leak and spill detection are desirable. The sensor consisted of two electrodes connected by interdigitated fingers (or electrodes, IDE) with nanotubes bridging the finger gaps. The IDE was fabricated using conventional photolithography with a nominal finger width of 10 μm and a gap size of 8 μm. The fingers were made of thermally evaporated Ti (20 nm) and Au (40 nm) on a layer of SiO_2 thermally grown on a silicon substrate. The sensing material was bulk-produced SWCNTs that were purified to remove amorphous carbon and metal impurities. Dispersal of the purified nanotubes was done in dimethyl formamide, $(CH_3)_2NC(O)H$, to create a suspension. The polymer coatings were chlorosulfonated polyethylene (CSPE), a synthetic rubber, and hydroxypropyl cellulose (HPC), a derivative of cellulose with both water solubility and organic solubility.

3. Zhang et al. (2006) functionalized SWCNTs by an electrochemical process with polyaniline (PANI). PANI is a versatile conducting polymer of the semi-flexible rod polymer family easily processed by melt or solution process, and is environmentally and thermally

stable. The PANI-SWCNT composite behavior was tested in NH_3, exhibiting detection limit of 50 ppb. The response time at room temperature was ~ minutes and the recovery time was a few hours.

4. Padigi et al. (2007) functionalized MWCNT with thiol (an organosulfur compound) for the detection of the first four fundamental aliphatic hydrocarbons: CH_3OH, C_2H_5OH, C_3H_7OH, and C_4H_9OH. High degrees of selectivity and sensitivity up to a detection concentration of 1 ppm were shown. In the presence of a chemical species, there was a shift of the resonant frequency peak.

In an attempt to decrease the working temperature of some materials that can be used as the sensing part and to increase the sensitivity of the devices, several studies have combined CNTs with oxides or metal particles. In these studies, the functionalization is not covalent, since the CNT structure is not modified: The process incorporates NPs or films on the surface of the nanotubes, thus changing the adsorption mechanism of the analytes with respect to the unmodified nanotubes. CNTs, then, play the role of the transducer in this type of sensors.

9.5.4 CNT-Based FET-Type Sensor

Several gas-sensing devices based on FET architecture have been proposed. *What is the principle of CNT FET sensors?* As may be recalled from the viewpoint of electrical conductivity, nanotubes can be metals or semiconductors, depending on the chirality of their structure. Semiconducting nanotubes are used to construct devices such as FETs. These devices are designed to detect analytes by means of a change in the electrical conductivity of the CNTs caused by the effect of the analyte on their surface. This effect can have two consequences: (i) they may promote a charge transfer from the analyte to the nanotube or (ii) the analyte may act as a scattering potential (i.e., potential causing the scattering of a particle). The charge transfer increases the conductivity when the analyte adsorbed is an electron attractor (hole donor). If the analyte is an electron donor, the number of holes on the nanotubes will decrease, which leads to a decrease in the electrical conductivity. This effect can be used to modulate FET-type devices with a third gate electrode. On the other hand, if the analyte acts as a scattering center, the current will decrease without shifting the characteristics.

What CNT FET gas sensors have been fabricated? Wong et al. (2003) used MWCNTs obtained by microwave plasma-enhanced chemical vapor deposition (MPECVD) on silicon substrates to construct nano FETs. The top part of the MWCNTs was coated with a sputter-deposited thin layer of palladium (100 nm), which also served as the gate electrode. The Pd layer acts as a catalytic agent enhancing the dissociation of the H_2 molecules into its atoms. The atomic hydrogen, adsorbed on the Pd NPs, transferred electrons to the CNTs. This decreased the number of hole carriers and, therefore, the conductivity

of the device. The researchers monitored the change in the electric current by substituting air for a flow of H_2 and observed the effect of temperature. They noted that there was a remarkable increase in sensitivity at higher operating temperatures (180°C).

An NO_x and NH_3 SWCNTs-based chemical sensor was fabricated by dielectrophoretic process (phenomenon in which a force is exerted on a dielectric particle when it is subjected to a nonuniform electric field; this force does not require the particle to be charged) on an FET structure (Lucci et al. 2006). It was found that NH_3 reduced the conductivity because of the charge transfer to the SWCNT, whereas NO_x induced an opposite effect.

9.5.5 MWCNTs/SnO$_2$ Ammonia Sensor

Hieu et al. (2008) developed an SnO_2/MWCNTs composite-based NH_3 sensor working at room temperature. The fabrication process involved the dispersion of MWCNTs in the SnO_2 dispersion. *What is the reason for combining tin oxide with CNTs?* It was intended to derive the merits of both materials in one device, i.e., blending two types of properties in one device. The SnO_2-based gas sensors are able to detect NH_3 gas with good sensitivity and response recovery time, but they only operate at elevated temperatures. In contrast, the CNT-based sensors can detect NH_3 gas at room temperature, but their sensitivity is low and response recovery time is excessively long. Thus the former is unable to detect NH_3 gas at room temperature and the latter has very long recovery and response times in detection of NH_3 gas at room temperature. But the composite MWCNTs/SnO_2 thin-film gas sensor amalgamates the desirable characteristics of both types of sensors, providing a high response and good recovery in detection of NH_3 gas at room temperature. It thus solves the problems of individual SnO_2-based and CNT-based sensors.

MWCNTs/SnO_2 composite sensor behaved as a P-type semiconductor. *What is the reason for this P-type character?* A plausible explanation is that the composite has a much higher CNT content; therefore, the major charge carriers are the holes, which are mainly supplied by CNTs.

At room temperature, the optimal composite sensor exhibited much higher response and faster response-recovery (less than 5 min) to NH_3 gas of concentrations ranging from 60 to 800 ppm, in comparison with the CNT-based NH_3 sensor.

9.5.6 CNT-Based Acoustic Gas Sensor

How do mechanical gas nanosensors work? Penza et al. (2004) demonstrated the integration of SWCNTs on to quartz-crystal microbalance (QCM) for alcohol detection at room temperature (Figure 9.11). The operating principle of QCM sensors is based on the fact that the change in the resonant frequency of a vibrating quartz-crystal resonator is proportional to the mass of the adsorbed gas molecules into a sensitive overcoating. A conventional 10 MHz QCM was
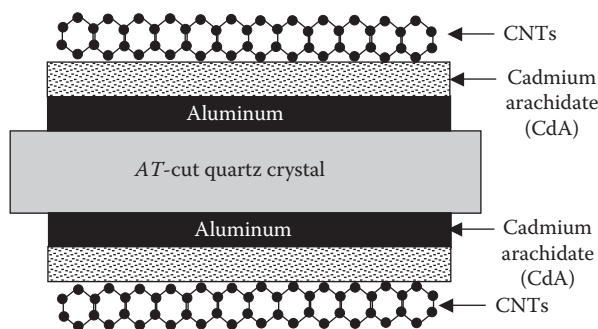
FIGURE 9.11
Quartz-crystal microbalance with linker-buffer layer of cadmium arachidate sensors to promote the adhesion of the single-walled carbon nanotubes as sensing overlayer on the sensor. (After Penza, M. et al., *Appl. Phys. Lett.,* 85, 2379, 2004.)

used as an acoustic sensor. It consists of a circularly shaped AT-cut quartz crystal with a diameter of 10 mm and a thickness of 0.1 mm. The aluminum electrodes deposited on both sides of the quartz were 100 nm thick and 4 mm in diameter. The active area for deposition of sensing CNTs was 12.5 mm^2. The noise of the uncoated 10 MHz QCM sensor was 0.5 Hz in 10 min.

The cadmium arachidate (CdA), prepared by the Langmuir–Blodgett (LB) technique, was used as multilayered buffer material on acoustic sensor to promote the adhesion of the SWCNTs as sensing overlayers on the sensors. The Langmuir–Blodgett technique is a method of depositing crystalline films one molecular layer at a time, by dipping the substrate into water containing a polymer that forms a single layer of molecular chains on the surface. This layer is then transferred from the water to the substrate. The dipping is repeated to create an ordered multilayer film that does not require poling to orient the molecules. An LB film is a set of monolayers, or layers of organic material one molecule thick, deposited on a solid substrate. It can consist of a single layer or many, up to a depth of several visible-light wavelengths.

The CdA was chosen as a linker-buffer material with hydrophobic surface characteristics; literally, hydrophobe means something with a fear of water. LB films consisting of tangled bundles of SWCNTs without surfactant molecules were successfully transferred on QCM.

The resonant frequency was the sensor output. The frequency decreases with respect to nitrogen reference ambient upon ethanol (ethyl alcohol, C_2H_5OH) exposure with a shift of 230 Hz with the injection of 59 mm Hg ethanol for the SWCNT-coated QCM. The resonant frequency decrease was ascribed to the mass of the alcohol molecules adsorbed into nanosensing SWCNTs. The frequency shift of the SWCNT-coated QCM was extremely higher than that of CdA-coated QCM with good reversibility and excellent repeatability: The response increased from 60 to 230 Hz for exposure to 59 mm Hg ethanol. Hence, the influence of the CNTs on alcohol sensitivity

was clearly discernible. This enhanced sensitivity could be accredited to typical high surface area of SWCNTs.

Example 9.2

The frequency of a quartz crystal is 10 MHz. Its active area is 10 mm². The frequency changes by 50 Hz on adsorption of a gas in the gas-sensitive film deposited on the active area of the sensor. Find the change in mass of the crystal upon gas adsorption.

Here, $A = 10\,mm^2 = 10 \times 10^{-2}\,cm^2 = 0.1\,cm^2$, $f_0 = 10\,MHz = 10 \times 10^6\,Hz = 10^7$ Hz, and $\Delta f = 50\,Hz$. Recalling the Sauerbrey equation from Equation 4.1

$$\Delta f = \frac{-2f_o^2 \Delta m}{A\sqrt{\rho_q \mu_q}} \tag{9.11}$$

where
$\rho_q = 2.648\,g\,cm^{-3}$
$\mu_q = 2.947 \times 10^{11}\,g\,cm^{-1} \times s^{-2}$

This equation is rearranged as

$$\Delta m = \frac{-\Delta f\left(A\sqrt{\rho_q \mu_q}\right)}{2f_o^2} = -\frac{50 \times 0.1 \times \sqrt{2.648 \times 2.947 \times 10^{11}}}{2 \times \left(10^7\right)^2}$$

$$= 2.2 \times 10^{-8}\,g = \frac{2.2 \times 10^{-8}}{10^{-9}}\,ng = -22\,ng \tag{9.12}$$

This is the mass of the gas responsible for increasing the mass of the crystal. The minus sign in this equation explains that Δf decreases as Δm increases.

9.6 Porous Silicon-Based Gas Sensor

What is the difference between crystalline silicon and porous silicon (PS) in terms of gas response? The physical properties of crystalline silicon are insensitive to the environment, but when crystalline silicon is reduced to the nanoscale, the resulting porous nanocrystalline structure strongly reacts, even explosively, with various analytes (Plessis 2008). The silicon Bohr exciton radius is about 5 nm and upon electrochemical engraving, a PS structure with wall thickness in the range of 2–5 nm is formed. The consequential nanocrystallites show a high chemical reactivity toward the environment.

The instability of the native surface termination is the major obstruction prohibiting pervasive applications of PS in the sensor field. Massera et al. (2004) found that the electrical performance of a PS NO_2 gas sensor was greatly improved by stabilizing its surface with long-standing exposure to high NO_2 concentrations prior to usage.

PS is sometimes used as a large surface area substrate on which NPs are deposited. *How is a PS-based hydrogen gas sensor made?* Luongo et al. (2005) developed a hydrogen sensor in which the sensing part consists of NPs of Pd deposited on a P-type silicon substrate. Electrochemical etching of the substrate was performed in HF/ethanol/H_2O, forming nanopores having average diameter ~10 nm on the surface. Subsequently, approximately 4 nm thick Pd film was deposited and stabilized by annealing at 900°C for 1 h, during which SiO_2 and PdO were formed. This process augmented the contact surface area of the sensor. Also, the electrical resistance increased because vacancies were produced in Si. The PS–Pd sensor was tested with a four-point probe configuration. The conductivity of the device was found to increase under nitrogen and with hydrogen concentrations of 0%–1.5%. The Pd acted as the catalyst for the dissociation of hydrogen, which transferred electrons to the silicon substrate. This relocation of electrons decreased the resistance. Since the PdO formed during the annealing was reduced at temperatures of 100°C or above, thereby degrading device performance, the sensor was operated at ambient temperature.

9.7 Thin Organic Polymer Film–Based Gas Sensors

How have polymer films contributed in gas sensing? Numerous sensors have been developed using few tens/hundreds of nanometer thick polymeric materials. Nohria et al. (2006) constructed a humidity sensor by depositing thin films of poly(anilinesulfonic acid) (SPANI) on a substrate by spin coating, from a solution of the polymer leading to a film 90 nm thick. As a second approach, films were also deposited by the layer-by-layer (LbL) nano-assembly technique, which negatively charges the substrate to couple the layers of a polycation [poly(allylamine hydrochloride)] (PAH), $(C_3H_7N)_n \cdot x$HCl and a polyanion [poly(styrene sulfonate)] (PSS). Once the layers had been deposited, the PSS was replaced by the SPANI to give a 26 nm thick film. On exposure to air, resistance of the sensor decreased as the relative humidity (RH) increased, with response times of 15 s for an LbL film sensor and 27 s for a spin-coated sensor.

Zhang et al. (2006) worked on the detection of vapors of tetrahydrofuran, C_4H_8O, clear, highly flammable liquid used as a solvent for natural and synthetic resins and in the production of nylon; benzene (C_6H_6), a colorless, flammable, aromatic hydrocarbon; toluene, C_7H_8 or $C_6H_5CH_3$, a clear,

water-insoluble liquid with the typical smell of paint thinners; cyclohexane, a cycloalkane with the molecular formula C_6H_{12}; *n*-hexane (petroleum naphtha), C_6H_{14}, a colorless liquid with a slightly disagreeable odor; carbon tetrachloride, also known as tetrachloromethane, CCl_4, a colorless liquid with a sweet smell; chloroform, $CHCl_3$, a clear volatile liquid with a strong smell like ether; ethyl acetate (ethyl ethanoate), EtOAc or EA, $C_4H_8O_2$, a colorless liquid with a fruity odor; diethyl ether, $(C_2H_5)_2O$, a colorless and highly flammable liquid; acetone, C_3H_6O, a colorless, volatile, extremely flammable liquid ketone; methanol, CH_3OH, a colorless, toxic, flammable liquid, used as an antifreeze, a general solvent and a fuel; ethanol, C_2H_6O, a clear, colorless liquid with a characteristic, agreeable odor; propanol, $CH_3CH_2CH_2OH$, a clear colorless liquid used as a solvent and as an antiseptic; and isopropanol, $(CH_3)_2CHOH$, a flammable liquid. They used a polystyrene (a vinyl polymer: $(C_8H_8)_n$) film with carbon nanofibers (cylinders consisting of tightly rolled graphite sheets that have diameters of 70–200 nm and lengths of 1–100 mm) to obtain stable sensors. The vapors were noticed through the change in electrical resistance shown by the device. The fact that the material is in the form of fibers gave higher stability because, when the gas was adsorbed, the NPs increased in size and had a propensity to join together.

What are the benefits of using polymers in gas sensing? (i) Organic polymers are cheap and highly sensitive to several analytes. (ii) The deposition techniques used to obtain the thin films are also relatively simple and economical. For these reasons, polymers offer a faster method of building gas sensors. *What are the problems associated with polymers?* The main problems associated with these materials are their (i) lack of selectivity and (ii) their lower stability regarding time and temperature. Additive materials such as carbon NPs or fibers have been explored to increase the selectivity and stability of polymer-based sensors.

9.8 Electrospun Polymer Nanofibers as Humidity Sensors

Can polymer nanofibers be used for moisture measurement? Yes. Li et al. (2009) prepared nanofibers of a polymer composite of a silicon-containing polyelectrolyte and soluble polyaniline (PANI: a unique type of polymer because it can be configured to conduct across a wide range, from being utterly nonconductive for insulation use to highly conductive for other electrical purposes) by electrospinning (ES). They investigated the humidity sensing properties of these polymers. ES is a simple and versatile technique for the easy preparation of continuous nanofibers.

The above nanofibers having a diameter of 250–500 nm formed a nonwoven carpet with a highly porous structure. The polymer composite nanofibers showed impedance change from 6.3×10^6 to 2.5×10^4 Ω with the rise of

RH from 22% to 97% at room temperature, exhibiting high sensitivity and good linearity on a semi-logarithmic scale. In addition, they showed fast and highly reversible response characterized by very small hysteresis of ~2% RH and short response time ($t_{90\%}$: 7 and 19 s for adsorption and desorption between 33% and 97% RH, respectively). The high resistance of electrospun nanofibers endorses their large porosity.

How is the humidity sensitivity of the sensor increased and what is the reason for this improvement? The modification of the underlying gold electrode with poly(diallyldimethylammonium chloride) (PDDA) prior to the deposition of nanofibers improved the humidity response of the sensor, due to the enhancement of contact between the nanofibers and underlying substrate. The introduction of PANI into the composite effectively decreased the impedance of the nanofibers.

9.9 Toward Large Nanosensor Arrays and Nanoelectronic Nose

What is odor made of? Odor is composed of molecules, each having a specific size and shape. All these molecules have a correspondingly sized and shaped receptor in the human nose. When a specific receptor receives a molecule, it transmits a signal to the brain and the brain identifies the smell associated with that particular molecule.

How does electronic nose (e-nose) mimic human nose? E-noses based on the biological model work in a similar manner to the human nose, albeit substituting sensors for the receptors, and transmitting the signal to a program for processing, in place of the brain. An e-nose is a device, which comprises an array of electronic chemical sensors with partial specificity (whether in the form of MOX semiconductors or via the use of different types of polymers) and an appropriate pattern recognition system, capable of recognizing simple or complex odors. Each sensor responds slightly differently to a given odor, and altogether, the sensors give an "odor fingerprint," a characteristic response pattern, such as a series of colors to each odor.

What are the common applications of the e-nose? The most common use for the e-nose at the present time is in the food and drink industries. In addition to this field, e-nose can be used in other areas such as classification and degradation studies of olive oils, qualitative and quantitative analysis of petroleum, detection of explosives, development of a field-odor detector for environmental applications, and quality control applications in the automotive industry.

How is the e-nose realized? Qi et al. (2003) showed that large arrays of devices containing multiple SWCNTs bridging electrodes can be easily produced with 100% yield by chemical vapor deposition (CVD) growth of nanotubes from a micropatterned catalyst. The ensemble of SWCNTs in each device

collectively exhibits substantial electrical conductance changes to electrostatic gating, owing to the high percentage of semiconducting nanotubes grown by CVD. The multiple tube devices (MT devices) are highly sensitive to chemical gating effects and, decisively, exhibit lower electrical noise than individual semiconducting SWCNT devices. Large arrays of nanotube chemical sensors with excellent characteristics are thus available. Further, Qi et al. (2003) carried out polymer functionalization of the MT devices to afford ultrahigh sensitivity for NO_2 detection and impart selectivity to nanotube sensors. They demonstrated that microspotting allows for functionalization of nanotube sensor arrays in a multiplex fashion, enabling detection of molecules in gas mixtures.

They investigated functionalization of nanotubes in the MT devices by various polymers for enhancing sensitivity and imparting selectivity to nanotube sensors, e.g., they adsorbed polyethyleneimine (PEI), $(C_2H_5N)_n$: linear form, onto SWCNTs in MT devices by simple immersion in a 20 wt% PEI/methanol solution for 2 h, rinsed with methanol, and then baked at 50°C for 1 h. The PEI-coated MT devices evolved into N-type due to electron transfer doping by high-density amine groups (a nitrogen atom with three bonds one of which is to a carbon atom) on PEI, similar to that observed with individual semiconducting SWCNTs. The PEI functionalized N-type MT devices were ultra sensitive to NO_2, responding to as low as 100 ppt (parts per trillion) of NO_2 (versus ~10–50 ppb for as-grown MT devices) with an appreciable conductance decrease. The lowest concentrations of NO_2 reliably detected by the N-type MT devices were several ppb with a response time of ~1–2 min (defined as the time for 80% conductance change to take place). The conductance change versus NO_2 concentration showed a linear dependence for $[NO_2] = 100$ ppt to 3 ppb and a tendency of saturation for higher $[NO_2]$.

They also explored selective detection of NH_3 with the MT devices and deciphered that coating an as-grown MT device with Nafion ($C_7HF_{13}O_5S \cdot C_2F_4$) blocks certain types of molecules from reaching nanotubes, including NO_2. This allows for NH_3 detection more selectively. Nafion is a sulfonated tetrafluoroethylene–based fluoropolymer-copolymer with a teflon [PTFE: $(C_2F_4)_n$] backbone and sulfonic acid (the class of organosulfur compounds with the general formula $RS(=O)_2$–OH, where R is an alkyl or aryl) side groups, well known to be perm-selective to –OH-containing molecules including NH_3 that tend to react with H_2O in the environment to form NH_4OH.

Further, they used microspotting to coat MT devices in an array with different polymers, aimed at detecting NO_2 and NH_3 in gas mixtures with multiplex-functionalized nanotube sensors. Microspotting of polymer solutions was done with a commercial micropin mounted on a micromanipulator equipped with an optical microscope that allowed for positioning the pin over the MT devices. It was noted that the conductance of a Nafion-coated device (spotted with 1% Nafion in a water droplet) decreases (due to NH_3 electron donation to the P-type device reducing the majority hole carriers),

while a PEI-coated (0.1% in the water droplet) device on the same chip shows no response to NH_3. When introducing 1 ppm of NO_2 into the environment with the presence of 500 ppm of NH_3, the Nafion-coated device gives no response while the PEI-coated device shows a significant conductance decrease. These results undoubtedly demonstrate that multiplexed nanotube sensor arrays are promising for detecting multiple molecules in complex chemical environments.

Chen et al. (2007) reported a brand-new platform, which can serve as an e-nose with good discrimination factors, constructed with four different semiconducting nanostructured materials including In_2O_3 NWs, SnO_2 NWs, ZnO NWs, and SWCNTs. The response of these sensors to hydrogen, ethanol, and NO_2 were measured at different concentrations and at both room temperature and 200°C.

Why is the integration of NWs with CNTs helpful in making e-nose? MOX NWs and SWCNTs have conspicuous differences in sensing mechanisms. In_2O_3, SnO_2, and ZnO NWs are N-type semiconductors, while SWCNTs are usually P-type semiconductors. Furthermore, the MOX NWs and nanotubes are believed to have different redox responses upon exposure to chemicals. This integration of both semiconducting MOX NWs and semiconducting SWCNT sensors therefore provides an important discrimination factor for improved selectivity. Besides, the integrated micromachined hotplate enables to control temperature individually and precisely for each sensor, with additional advantages such as ultralow power consumption (~60 mW) and short response time (20°C–300°C within 1 min). This temperature control works as the second discrimination factor. With these two discrimination factors, they achieved a "smell-print" library by detecting important industrial gases such as hydrogen, ethanol, and nitrogen dioxide. In addition to this, they used principal component analysis for data processing. Principal component analysis of the sensing results showed great discrimination of those three tested chemicals. Thorough analysis revealed clear improvement of selectivity by the integration of CNT sensors. This nanoelectronic nose approach has immense potential for detecting and discriminating between a wide variety of gases, including explosive ones and nerve agents.

9.10 CNT-, Nanowire-, and Nanobelt-Based Chemical Nanosensors

An FET sensor has the structure of a common three-electrode transistor, where the source and drain electrodes bridge the semiconductor channel, and the gate electrode modulates the channel conductance. In the case of FET nanosensors, the semiconductor channel is made of a nanomaterial and is used as the "sensing" component of the device. Semiconductor channels

can be fabricated using several nanomaterials, including CNTs and NWs. In order to provide selectivity toward a unique analyte, a specific recognition group (also called a receptor, ligand, or probe) is anchored to the surface of the semiconductor channel. This receptor is typically chosen to recognize its target molecule (also called analyte) with a high degree of specificity and affinity.

9.10.1 CNT-Based ISFET for Nano pH Sensor

CNTs come in two different forms: MWCNTs and SWCNTs, which range in diameter from 1 to 10 and 10 to 50 nm, respectively. About 70%–80% of SWCNTs display semiconducting properties, whereas 70%–80% of MWCNTs show metallic properties.

How is a CNT handled for its accurate and exact positioning between source and drain? Dielectrophoresis (DEP) is a process through which neutral particles, such as CNTs, are translated through a suspending medium in a nonuniform electric field, which is generated between a pair of electrodes. DEP is used to separate, trap, and sort cells, bacteria, etc. DEP technique is an accepted, preferred method for aligning CNTs. Dong and Wejinya (2009), and Dong et al. (2009) carried out alignment and testing of CNT *I–V* characteristic to verify if they are metallic or semiconducting. CNTs with metallic properties serve as NWs taking the place of the inversion layer in ISFET, while the other kind of CNTs with semiconducting properties are used for the fabrication of nanotransistors. The microchip is originally made up of four layers from the bottom up: silicon wafer, 300 Å of silicon dioxide, 200 Å of chromium, and 3000 Å of gold. It has triangular electrodes with an angle of 30°, and having electrode gaps ranging from 2 to 30 μm. The stock solution is a mixture of MWCNTs, surfactant and deionized (DI) water. A sonicate process is necessary to suspend CNTs uniformly. An AC power supply is used to produce the DEP forces. For this purpose, a function generator (electrical waveform generator) is used with an oscilloscope (electronic test instrument that allows observation of constantly varying signal voltages) to monitor the power function generated. For testing *I–V* characteristics, the gap distance is about 30 μm, and a micropipette (a very small pipette used in microinjection) is used to deposit a 1.5 μL CNT droplet from the solution onto the gap. The function generator is turned on and DEP forces are formed between the electrodes. After CNT alignment has been successfully done, a NW consisting of MWCNTs is formed, connecting the source and drain electrodes. Its resistance measured by a multimeter, is ~1 kΩ. For *I–V* measurement, a DC power generator is used to supply DC potential and the corresponding current values are measured by the multimeter. SEM is adopted to examine how the CNT alignment is done.

Does the current sensing AFM help in CNT alignment and I–V measurement? Yes, the above process of CNT alignment and *I–V* measurement is also done by current sensing AFM (CSAFM) function, using which many experiments

related to electrical properties of nanostructures are performed, including impedance measurement and testing of *I–V* characteristic. Here an ultra-sharp AFM cantilever, coated with conductive film, probes the conductivity and topography (shapes and features) of the sample surface concurrently. A bias voltage is applied to the sample while the cantilever is kept as virtual ground (a node of the circuit that is maintained at a steady reference potential, without being connected directly to the earth). During scanning, the tip force is held constant and the current is used to construct the conductivity image of the surface. From these experiments, a conclusion on whether these SWCNTs are semiconducting is established.

9.10.2 NW Nanosensor for pH Detection

Cui et al. (2001) prepared single-crystal boron-doped (P-type) Si NWs by a nanocluster-mediated vapor–liquid–solid growth method. Devices were fabricated by flow aligning Si NWs on oxidized silicon substrates and then making electrical contacts to the NW ends with electron-beam lithography. *How is the silicon dioxide surface modified to enable pH measurements?* The silicon NW solid state FET, whose conductance is modulated by an applied gate, was transformed into a nanosensor by transforming the silicon oxide surface with 3-aminopropyltriethoxysilane (APTES) to provide a surface that undergoes protonation and deprotonation, so that changes in the surface charge can chemically gate the Si NW. Measurements of conductance as a function of time and solution pH demonstrated that the NW conductance increases stepwise with discrete changes in pH from 2 to 9 and that the conductance is constant for a given pH; the changes in conductance are also reversible for increasing and/or decreasing pH.

How are the results interpreted? These results are understood by considering the mixed surface functionality of the modified Si NWs (Figure 9.12). Covalently linking APTES, 3-aminopropyl)triethoxysilane ($C_9H_{23}NO_3Si$) to Si NW oxide surface results in a surface terminating in both $-NH_2$ and $-SiOH$ groups, which have different dissociation constants, pK_a. At low pH, the $-NH_2$ group is protonated to $-NH_3-$ and acts as a positive gate, which depletes hole carriers in the P-type Si NW and decreases the conductance. At high pH, $-SiOH$ is deprotonated to $-SiO-$, which correspondingly causes an increase in conductance. The observed linear response is ascribed to an approximately linear change in the total surface charge density (versus pH) because of the united acid and base behavior of both surface groups. To uphold this contention, Cui et al. (2001) also carried out pH-dependent measurements on unmodified (only-SiOH functionality) Si NWs. These conductance measurements showed a nonlinear pH dependence: The conductance change is small at low pH (2–6) but large at high pH range (6–9). Notably, these pH measurements on unmodified Si NWs were in conformity with previous measurements of the pH-dependent surface charge.
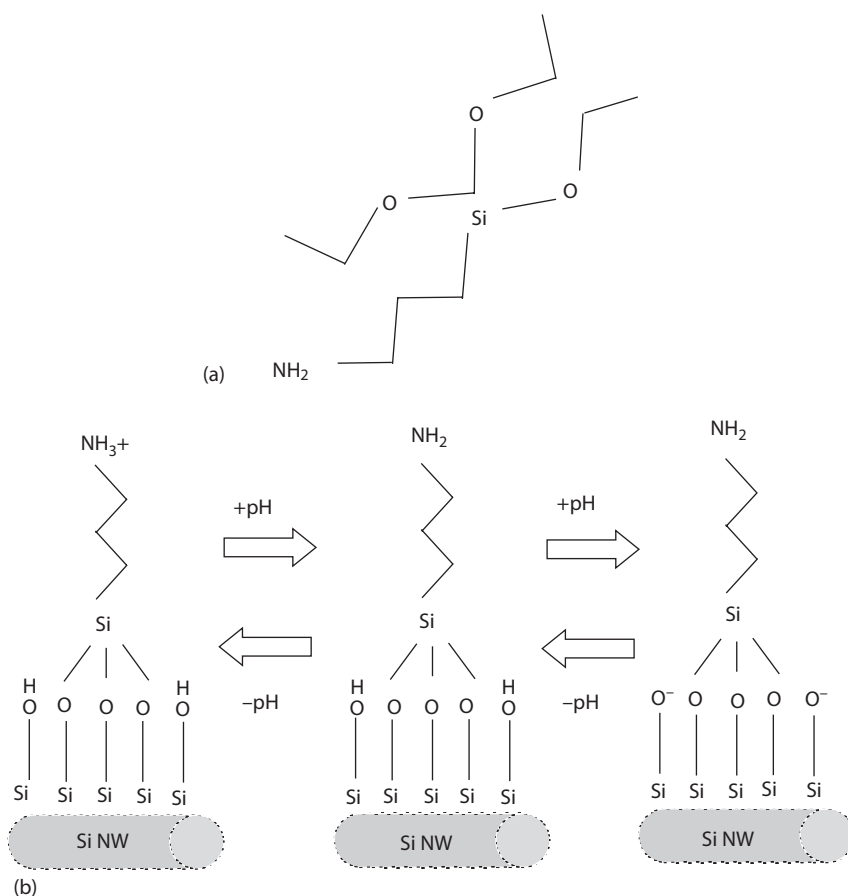
FIGURE 9.12
(a) Structural formula of 3-aminopropyltriethoxysilane (APTES). (b) APTES-modified Si NW surface illustrating changes in the surface charge state with pH. (After Cui, Y. et al., *Science*, 293, 1289, 2001.)

Hsu et al. (2005) synthesized Si NWs on silicon substrates via a catalytic reaction in N_2 atmosphere by thermal CVD at 955°C. The Si NWs have an average diameter of approximately 30–50 nm and a length of up to a few tens of micrometers. Silicon bulk materials and Si NWs were used alternately as the sensing layer in an extended-gate FET configuration for measuring solution pH. Experimental results showed that the pH sensitivity of silicon bulk materials is poor. However, good pH sensing properties of Si NWs, with a sensitivity of 58.3 mV pH^{-1}, was observed. Therefore, it was affirmed that the pH sensitivity of silicon bulk materials was greatly improved by downsizing them to the nanoscale. The sensitivities approach those of the pH sensors made of MOXs, such as SiO_2/Si_3N_4, SnO_2, and Ta_2O_5, which have sensitivities of 54, 58, and 59.2 mV pH^{-1}, respectively. The boron-doped Si NWs oxide

surface resulted in a surface terminating in –SiOH groups, which would be deprotonated to –SiO⁻, causing an increase in the conductance of the FET.

According to Hsu et al. (2005), Si–O asymmetric stretching bonds act as sites for H^+ ion adsorption. A higher Si–O asymmetric stretching bond density exists on the surface of the Si NWs sensing layer, than on the Si bulk materials in the same magnitude sensing window accounting for improved pH sensitivity of NWs.

9.10.3 ZnS/Silica Nanocable FET pH Sensor

The NWs that have been extensively developed for sensing are those of silicon because of the gigantic knowledge that has existed for the chemical modification of native Si oxide surfaces. *Excluding the CNT, is there any alternative material that has been used as a NW in FET?* Yes, the FET of ZnS/silica (SiO_2) core/shell nanostructure is attractive to chemical sensors due to (i) the excellent and reproducible electronic characteristics of ZnS, (ii) the self-assembled silica shell is a natural insulator on the surface of the ZnS NW that serves as the protection layer against oxidation, and (iii) the colossal knowledge available about the chemical modification of silica surfaces.

He et al. (2007) demonstrated the reproducible transport characteristics of the ZnS/SiO_2 core/shell nanocable FETs, in which metal electrode and electrolyte solutions were used as a gate. The nanocable-based devices were used for label-free, real-time, and sensitive detection of chemical species. Amine- and hydroxyl (OH)-functionalized ZnS/silica nanocables exhibit linear pH-dependent conductance, which could be elucidated in terms of the changes of surface charge during protonation and deprotonation.

FET devices were fabricated using standard procedures with back gate geometry. The nanocables were synthesized using vapor–liquid–solid growth. For the fabrication of nanocable FETs, the electrode pattern was designed to have a few parallel electrodes separated by 5–20 μm employing standard photolithography and a lift-off process. The synthesized ZnS/silica nanocables were transferred from the Si substrate to pre-patterned Au/Ti electrodes by touching the nanocable sample with the electrodes. A single nanocable device could be achieved easily using this method. To firmly contact the Au/Ti metal electrodes and ZnS core of the nanocable, focused ion beam (FIB) microscopy was employed to cut the nanocable at the two ends so that the ZnS core was exposed, and then, a Pt mixture was deposited at the ends to make contacts between the ZnS core and Au/Ti electrodes. The oxygen plasma treatment was carried out to get rid of the surface contamination of the ZnS/silica nanocable device before chemical sensing.

How is the ZnS/silica nanocable FET used as a chemical sensor? Chemical sensing in solution was carried out by monitoring electrical conductance through the surface-modified ZnS/silica nanocable device during buffer solution additions with different pH values. Here, the ZnS/silica nanocable-based FET, whose conductance is modulated by an applied gate, is transformed
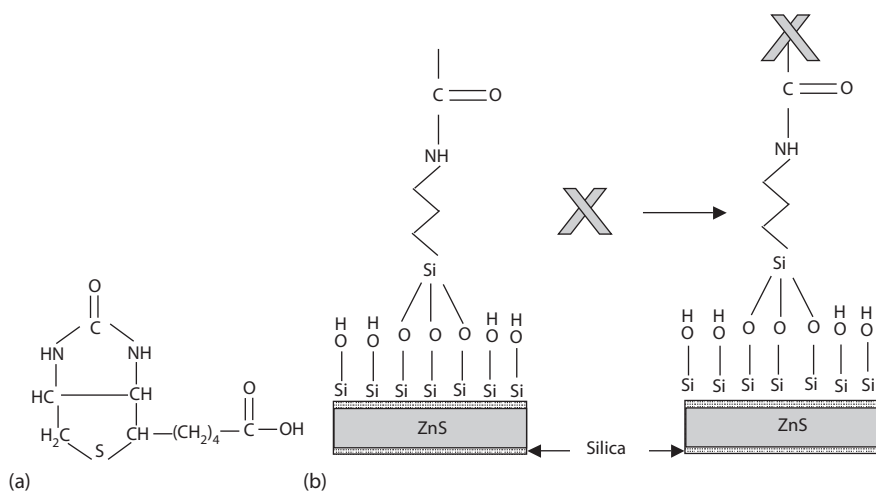
FIGURE 9.13
(a) Structure of biotin. (b) Biotin-modified ZnS/silica nanocable (left) and binding of streptavidin to the surface of the nanocable (right). (After He, Jr. H. et al., *J. Phys. Chem. C*, 111, 12152, 2007.)

into a chemical sensor by altering the surface of the silica shell with APTES to provide a surface that can undergo protonation and deprotonation, where changes in the surface charge are able to chemically gate the nanocable-based transistors. After the cleaning process in oxygen plasma to remove contaminants has been completed, ZnS/silica FETs were placed in the chamber under the saturation vapor pressure of a 1% ethanol solution of APTES overnight and then rinsed with ethanol thrice to achieve surface functionalization of nanocable FETs.

Measurements of conductance as a function of time and pH demonstrate that the NW conductance increases in steps corresponding to the pH values. The proposed mechanism for pH sensing is the same as described in the preceding section.

To explore NW-based sensors for selective biological recognition, He et al. (2007) functionalized the silica shell of a nanocable functionalized with biotin, $C_{10}H_{16}N_2O_3S$. The conductance-time measurement shows that the conductance of the biotin-modified nanocable decreases rapidly and considerably to a constant value upon addition of the steptavidin solution. The mechanism is shown in Figure 9.13.

9.10.4 Bridging Nanowire as Vapor Sensor

The high surface area and small volume of the NWs makes them especially attractive as effective field-effect sensing elements. In a field-effect sensor, charge from an adsorbed or nearby analyte induces compensating charge in the NW, modulating its conductance and, consequently, the

current flowing between two electrodes. The device, therefore, acts as a chemically sensitive FET, with adsorbed charged species acting as the "gate." NW sensors can be very sensitive because a significant fraction of the carriers in the NW can be depleted so that a small number of carriers induced by the analyte changes the conductance significantly. *What is the problem in interfacing NWs with circuitry?* Using the NWs requires that they be connected to electrodes that can interface with other circuitry. The size mismatch between the nanoscale wires and conventional electrodes has impeded their adoption.

How does the bridging NW structure help in securing good electrical connection with microelectrodes? Kamins et al. (2006) demonstrated the use of "bridging NW" structure (Figure 9.14) for sensing chemical species, such as NH_3 and HCl vapors. Metal-catalyzed, self-assembled silicon NWs were grown between two electrically isolated electrodes so that intimate contact was made between the NWs and microscale electrodes during growth. Because the connection was between Si NWs and heavily doped Si electrodes of the same conductivity type, good electrical connection was readily obtained. Connection between the Si electrodes and metal contact pads was achieved using conventional integrated-circuit techniques; the contact area was large so that metal-semiconductor contact resistance did not limit the current flow. The bridging NW structure is amenable to integration of the NW sensing elements with related electronics formed by conventional integrated-circuit technology.

Positive adsorbed charge depletes mobile holes from the region near the surface of the NW, decreasing its conducting area and conductance. Negative adsorbed charge attracts additional mobile holes to the surface region of the NW, increasing its conductance. For the P-type NWs used in these experiments, a current decrease corresponds to a positively charged species adsorbed on the NWs decreasing the density of positive mobile carriers (holes), while an increase pertains to a negative charge on the NWs
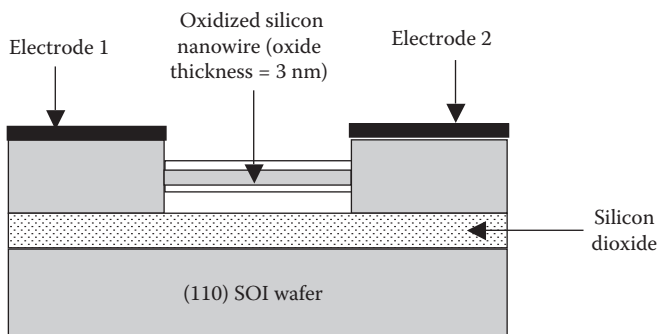
FIGURE 9.14
Briding nanowire structure with nanowire grown between two silicon electrodes electrically isolated from the substrate by the oxide of an SOI wafer. (After Kamins, T. I. et al., *Nanotechnol.*, 17, S291, doi: 10.1088/0957-4484/17/11/S11, 2006.)

increasing the density of holes. Thus, the decrease of current flowing through the NWs when exposed to NH_3 vapor arises from positive surface charge while the increase of current when exposed to HCl vapor is due to negative surface charge. The positive charge on the NW surface resulting from exposure to NH_3 vapor was undecidedly attributed to the reaction

$$NH_3 + H_2O \Leftrightarrow NH_4^+ + OH^- \tag{9.13}$$

while the negative charge from exposure to HCl vapor was attributed to Cl^-.

9.10.5 Palladium Functionalized Si NW H_2 Sensor

Chen et al. (2007) fabricated a hydrogen sensor by functionalizing Si NWs with Pd NPs. Recalling the Pd-functionalization process briefly (see Section 3.6.3), oxide sheath on as-grown 1D Si NWs synthesized by thermal evaporation of high purity SiO (99.9%) mixed with 1% Sn powder, was etched in a 5% HF aqueous solution, followed by immersion of the sample in a saturated palladium chloride ($PdCl_2$) solution. Pd^{2+} ions were reduced and Pd metal particles deposited on the surface of Si NWs. The Pd-modified Si NWs were dispersed on a silicon wafer with a 300 nm oxide layer. Gold electrodes were deposited on the sample by thermal evaporation through a shadow mask. *I–V* measurements were carried out with a two-probe measurement system.

The oxide-removed Si NWs showed N-type properties due to point defects and surface states in Si surfaces. The conductance of the Pd-coated Si NW was smaller than that of the pure Si NW due to electron depletion at the Pd and silicon interface. Exposure to 5% hydrogen increased the current signal of the sensor 20-fold. The effect of hydrogen on the conductivity was reversible. The response time was 3 s only, which is much faster than that of the macroscopic Pd wire sensor.

How does the H_2 gas sensor work? The mechanism of the gas sensor is expounded by electron depletion (or even inversion) at the Pd/Si interface and consequently reducing electrical conductivity. At room temperature and atmospheric pressure, palladium absorbs up to 900 times of its own volume of hydrogen. Upon exposure to hydrogen, Pd reacts with hydrogen to form hydride (PdH_x) rapidly. The Fermi level of PdH_x moves to a higher energy, which is above the dense region of the *d* bands; this brings up the Fermi level at the interface and results in the transition of electrons from Pd to Si NWs, returning them to a more N-type behavior. So depletion width is a critical parameter that controls how significantly the Pd Fermi level shift can modulate the NW conductance. Due to the reversible reaction between Pd and hydrogen, PdH_x releases hydrogen quickly and leads to a decrease in conductance again. Compared with the back gate modulation in FET with a partition of thick silicon oxide layer, the coated Pd NPs modulate the conductance of Si NWs directly and more effectively.

9.10.6 Polymer-Functionalized Piezoelectric-FET Humidity Nanosensor

Lao et al. (2007) demonstrated a humidity/chemical nanosensor based on piezoelectric field-effect transistor (PE-FET). *What is a PE-FET?* It is an FET structure with the gate insulator made of a piezoelectric material that develops a voltage across it upon deformation. The device (Figure 9.15) has a single-side coated ZnO NB gate functionalized with multiple layers of polymers using an electrostatic self-assembling process.

How is the stress produced in the ZnO NB? Polymers for functionalization of NB were anionically charged poly N-isopropylacrylamide (PNIPAM), $C_6H_{11}NO$, and poly(diallyldimethylammonium chloride) (PDADMAC), $(C_8H_{16}ClN)_n$. These polymers were closely packed before vapor exposure. Upon exposure to water vapor, the polymers underwent a hydration process. As a result, the volumes of these polymers were increased by several factors. Because the polymers were coated only at one side of the ZnO NB,
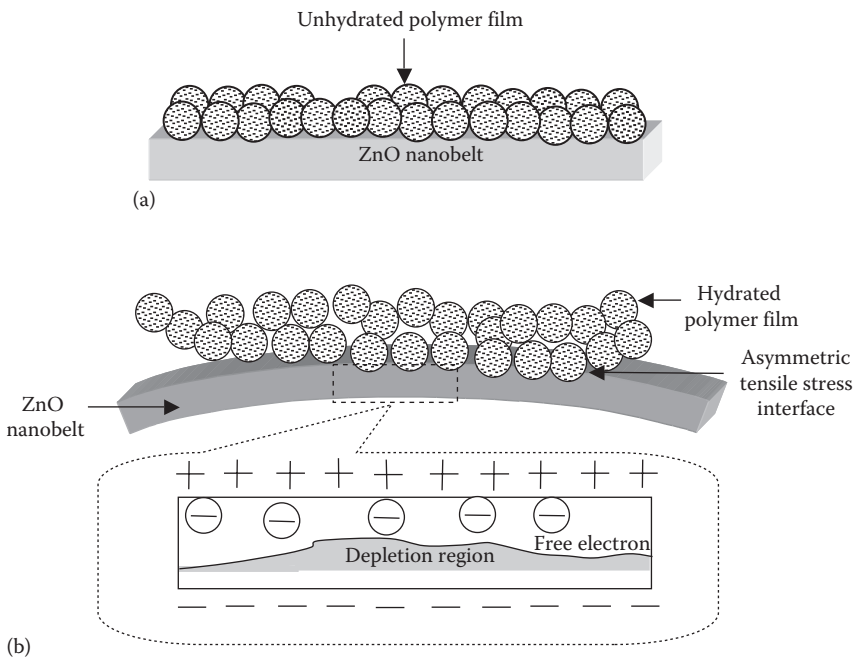
FIGURE 9.15
Polymer functionalized piezoelectric field effect transistor (PE-FET: (a) Dehydrated polymer film lies straight on zinc oxide nanobelt causing no deformation of the belt. (b) Upon hydration, the polymer swells and expands in volume causing bending of the ZnO belt with attendant piezoelectric field generation across the ZnO nanobelt. Positive and negative charges are produced, respectively, on the outer and inner surfaces of ZnO nanobelt. The positively charged top surface attracts free electrons while the repulsive force of the negatively charged bottom surface creates a depletion region adjacent to it. Dehydration in a dry environment results in reverting of the sensor to condition (a). (After Lao, C. S. et al., *Appl. Phys. Lett.*, 90, 262107-1, 2007.)

the swelling of polymers generated an asymmetric tensile stress at the contact surface with ZnO NB. Consequently, ZnO NB was bent, resulting in an asymmetric strain across the thickness of the NB. A bent ZnO NB could produce a piezoelectric potential across the NB due to the strain-induced piezoelectric effect. With the stretch and compression effects of a deformed ZnO NB, positively charged and negatively charged surfaces were produced at the outer and inner bending surfaces of a ZnO NB. Consequently, a piezoelectric field was built across the ZnO NB.

Thus the operation of the PE-FET relies on the self-contraction/expansion of the polymer, which builds up a strain in the piezoelectric NB and induces a potential drop across the NB that serves as the gate voltage for controlling the current flowing through the NB. In return, the deformation of ZnO NB produces a piezoelectric field across the NB, which serves as the gate for controlling the flow of current along the NB.

The device is a useful component for nanopiezotronics.

9.11 Optochemical Nanosensors

9.11.1 Low-Potential Quantum Dot ECL Sensor for Metal Ion

What are the attractive features of quantum dot (QD)-based ECL analysis method? The QD-based electrochemiluminescent (ECL) analytical technique has developed fast in many fields because of its advantages such as low cost, high sensitivity and stability, unsophisticated instrumentation, and applicability to a wide range of analytes.

Cheng et al. (2010) constructed the ECL sensing system for low-potential detection of metal ion. They immobilized surface-unpassivated cadmium telluride (CdTe) QDs on a glassy carbon electrode. The surface-unpassivated CdTe QDs were prepared using meso-2,3-dimercaptosuccinic acid (DMSA), with the molecular formula $C_4H_6O_4S_2$, as a stabilizer to cap CdTe QDs. DMSA is an organosulfur compound with the formula $HO_2CCH(SH)CH(SH)CO_2H$, containing two carboxylic acid and two thiol groups. The immobilized QDs showed a strong cathodic ECL emission peak at 0.87 V with an onset potential at 0.64 V (versus Ag/AgCl/ saturated KCl) in air-saturated, pH 9.0 HCl Tris [tris(hydroxymethyl)aminomethane: $C_4H_{11}NO_3$] buffer. On the basis of the competition of metal ion to the stabilizer, the quenching effect of metal ion on ECL emission was observed, leading to a responsive chemical sensing application. A simple analytical method for Cu^{2+} detection was developed utilizing the quenching effect on the cathodic ECL emission of DMSA-CdTe QDs. This followed the behavior of the fluorescence (FL) quenching principle described by the Stern–Volmer equation, an expression relating FL quenching to the concentration of the quenching substance:

$$\frac{I_0}{I} = 1 + K_{sv} \times [Q] \tag{9.14}$$

where
   I_0 is the initial ECL intensity
   I is the ECL intensity at a given concentration of quencher $[Q]$
   K_{sv} is the quenching constant

The Stern–Volmer relationship allows the exploration of the kinetics of a photophysical *intermolecular* deactivation process, such as FL, an *intramolecular* deactivation (quenching) process. An *intermolecular* deactivation occurs where the presence of a chemical species β accelerates the decay rate of a chemical α in its excited state.

The sensor showed a linear relation in the Cu^{2+} concentration range from 5.0 nM to 7.0 μM. The detection limit was 3.0 nM and the quenching constant was $4.6 \times 10^5 M^{-1}$.

After the detection procedure, the ECL emission of QD-modified GCE was not recovered to the original intensity prior to the quenching of Cu^{2+}. This pointed toward a structure destruction of the QDs by Cu^{2+}. The competitive binding of Cu^{2+}, with stronger metal-S interaction than the Cd–S bond, to the stabilizer led to a quenching effect on ECL emission. This was substantiated to be due to the thermodynamic tendency for formation of Cu-S on the DMSA-CdTe QD surface.

9.11.2 BSA-Activated CdTe QD Nanosensor for Sb³⁺ Ion

Why is it necessary to establish sensitive and accurate analytical methods for quantitative determination of antimony? It is necessary because antimony is a toxic element and the upper limit of antimony in domestic water is $0.005 mg L^{-1}$. Antimony toxicity takes place either through occupational exposure or during therapy. Exposure to high levels of antimony has a variety of adverse health effects. Ingesting or inhaling certain compounds of antimony has injurious effects upon body tissues and functions. Acidic fruit juices containing antimony oxide dissolved from the glaze of cheap enamelware containers have also caused antimony poisoning.

Ge et al. (2010) designed an antimony selective nanosensor based on bovine serum albumin (BSA)-fluorophore system, where the fluorophore is a thioglycolic acid (TGA)-capped CdTe QD nanoparticle with BSA as an antimony conjugate (Figure 9.16). TGA or mercaptoacetic acid is the organic compound $HSCH_2COOH$. It contains both a thiol (mercaptan) and a carboxylic acid. In BSA, amino groups are available for the conjugation with carboxylic acid, –COOH, group capping QD NPs via amide formation. BSA is able to couple to TGA-capped QDs through an amide linkage offering opportunity for N to QD charger-transfer. Hence, BSA conjugates to TGA-capped CdTe via an amide link interacting with carboxyl of the TGA-capped CdTe.
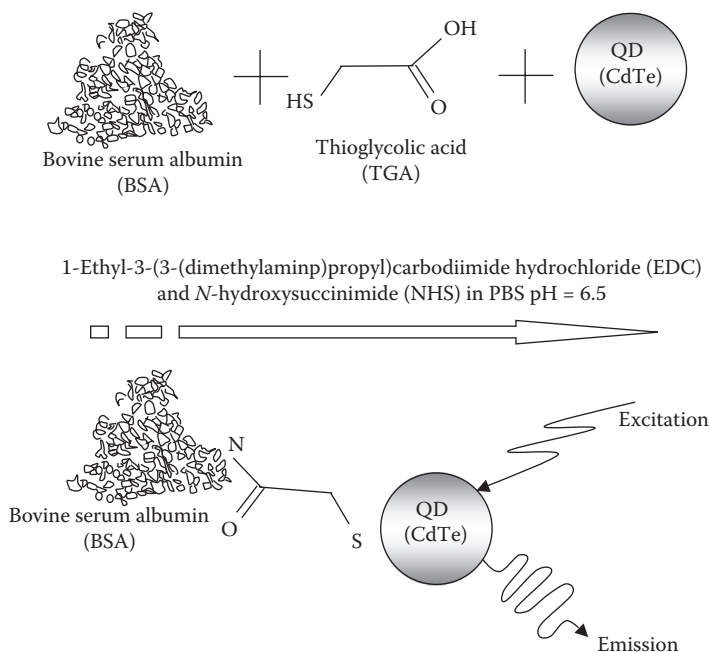
FIGURE 9.16
The covalent linking of BSA on the surface of TGA-capped QD nanoparticles by EDC/NHS coupling reaction. QD modification utilizes the property that BSA has amino groups available for the conjugation with carboxylic acid group capping QD nanoparticles via amide formation. The emission is switched on when the excitation light irradiates the TGA-QD-BSA conjugates. (After Ge, S. et al., *Analyst*, 135, 111, 2010.)

In the aforementioned work, hydrothermal route was selected for the synthesis of water-soluble TGA-capped CdTe QDs. If antimony is absent and the valence band of the QD is at higher energy than a molecular orbital on the ligand, the electron transfer from BSA to the QDs results in enhancement of QD intensity.

How is FL quenched when antimony is present? If the molecular orbital implicated is also involved in antimony binding, then in the presence of antimony, the energy level is no longer available. Due to its unavailability, electron migration is not permissible, and therefore it cannot occur. Thus the QD FL is quenched in the presence of antimony (Figure 9.17). When antimony ion enters the BSA, the lone pair electrons of the nitrogen and oxygen atom become involved in the coordination, switching off the QD emission and a dramatic quenching of the FL intensity results. Extremely low concentrations of antimony ions thus become detectable.

A good linearity between change in intensity (ΔI) and the concentration of antimony was observed in the range of 0.10–22.0 mg L^{-1} with a correlation coefficient of 0.9984, and the detection limit was 2.94×10^{-8} g L^{-1}. This was
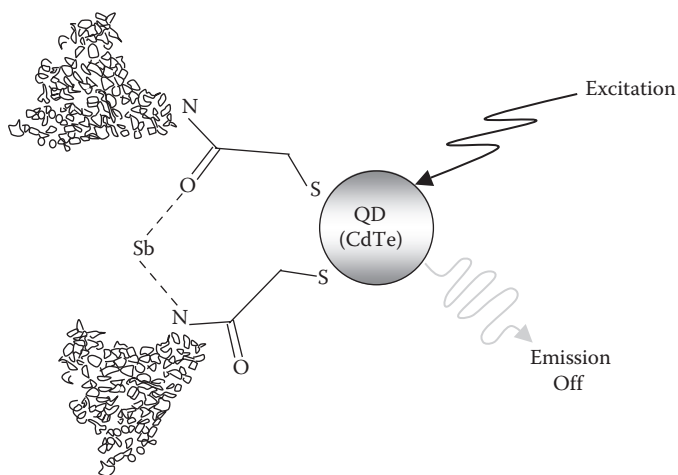
FIGURE 9.17
The quenching mechanism of TGA-CdTe-BSA-Sb. The quenching is indicated by representing the emission by a gray line meaning that the emission is turned off. When antimony ion enters the BSA, the lone electron pairs of nitrogen and oxygen atoms participate in the coordination, switching off the QD emission. Consequently, drastic quenching of the fluorescence intensity takes place enabling the detection of minute concentrations of Sb^{3+} ions. (After Ge, S. et al., *Analyst*, 135, 111, 2010.)

estimated as the concentration of analyte that produced an analytical signal equaling thrice the standard deviation of the background FL.

What is the effect of interfering ions? In a study of interferences, the antimony-sensitive TGA-QD-BSA sensor showed good selectivity. Even the presence of high concentrations of SO_4^{2-}, NO_3^-, Cl^-, F^-, K^+, Na^+, and NH_4^+ ions did not bring about any changes in the emission of the TGA-QD-BSA conjugates. This was probably due to poor complexation of the above ions with BSA that do not change the quenching of the QD; a complexation reaction is a chemical reaction that takes place between a metal ion and a molecular or ionic entity.

Thus interferences by metal ions in the detection were minimal. Therefore, a simple, fast, sensitive, and highly selective analysis for antimony was achieved. Furthermore, the sensor was applied successfully to the determination of antimony in real water samples with satisfactory results.

9.11.3 Functionalized CdSe/ZnS QD Nanosensor for Hg(II) Ion

What is the reason for notoriety of mercury? Mercury is a widespread, dangerous global pollutant. Mercury poisoning, also known as *mercuralism*, is the phenomenon of intoxication by contact with mercury. Mercury poisoning from amalgam dental fillings is the root cause of a multitude of difficult-to-diagnose and often life-threatening diseases.

On what principles is the QD-based Hg nanosensor designed? (i) The FL efficiency of QDs is sensitive to the presence and nature of adsorbates at the surface of QDs. Therefore, a chemical sensing system based on QDs can operate using FL changes induced by molecular recognition at the surface of QDs. (ii) L-carnitine (LC) is a water-soluble vitamin-like compound (vitamin BT), which modifies QDs via chelation of carboxyl group with heavy metal ions; chelation is the formation or presence of two or more separate bindings between a polydentate (multiple bonded) ligand and a single central atom. LC ($C_7H_{15}NO_3$) is a quaternary ammonium compound biosynthesized in the body from the amino acids lysine and methionine. It increases the use of fat as an energy source by transporting fatty acids into the mitochondria, where they are burned to release energy for body functions.

Li et al. (2008) prepared LC capped CdSe/ZnS core/shell QDs (Figure 9.18), which were capable of selectively responding as selective fluorescent probe for the determination of mercury ions. Monodisperse CdSe QDs were synthesized and were tri-*n*-octylphosphine oxide (TOPO) protected. Then, the freshly washed CdSe/ZnS/QDs were dispersed in the ethanol solution of
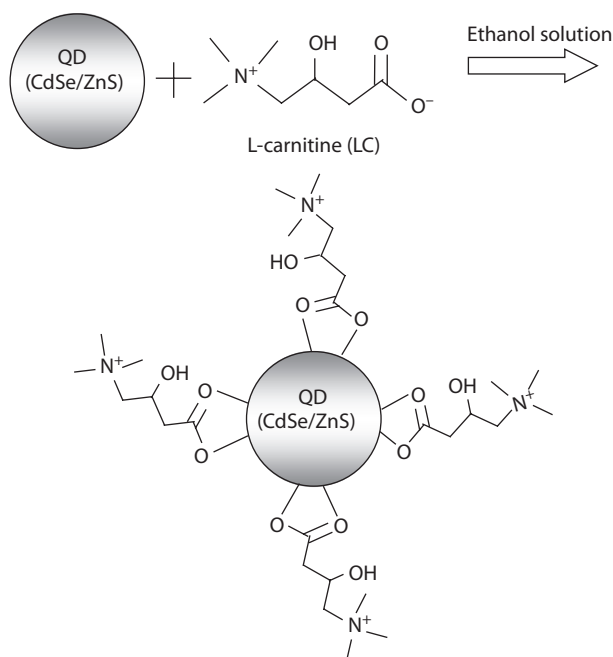
FIGURE 9.18
Synthesis of L-carnitine-capped QDs by the surface modification of QDs via carboxyl anchor group of L-carnitine for Hg(II) nanosensor. (After Li, H. et al., *Microchim. Acta*, 160, 119, doi: 10.1007/s00604-007-0816-x, 2008.)

LC. Aggregated QDs were removed by centrifuging (spinning about a central axis to separate contained materials of different specific gravities), and the target product of LC-capped QDs was separated.

Luminescence emission of the LC capped QDs was highly sensitive to the presence of mercury ions in ethanol. The fluorescent emission of LC-capped QDs was increasingly quenched by adding mercury ions. The result indicated that the quenching metal ions remained adsorbed or linked to the QD NPs.

The finding that mercury ions quenched the FL intensity of LC-capped QDs in a concentration dependence described by a Stern–Volmer-type equation, was applied to develop a method for the determination of mercury ions. The detection selectivity was checked by conducting the FL titration of LC-capped CdSe/ZnS/QDs with various metal ions. Only Hg(II) ions showed strong quenching of LC-capped QDs FL. The experimental data was understood in terms of strong affinity to Hg(II) ions with the surface ligands of QDs. The interference of alkali, alkali earth ions, Ni(II) ions, and Zn(II) ions was very frail. Thus quenching of the luminescence emitted by the synthesized NPs allowed the detection of mercury concentration as low as 0.18 mM, affording a very sensitive detection system for this toxic chemical species.

Example 9.3

Mercury ions quench the FL emission of LC-capped QDs in accordance with the Stern–Volmer equation

$$\frac{I_{max}}{I} = 1 + K_{sv} \times \left[Hg^{2+}\right] \qquad (9.15)$$

where

I_{max} is the luminescence intensity of LC-capped QD in a mercury ion-free solution

I is the luminescence intensity of LC-capped QD in a mercury ion solution

$K_{sv} = 1.167 \times 10^4 \, M^{-1}$

If the ratio I_{max}/I is observed to be 5, what is the mercury ion concentration $[Hg^{2+}]$ in the given solution?

From Equation 9.15

$$\left[Hg^{2+}\right] = \frac{\left(I_{max}/I\right) - 1}{K_{sv}} = \frac{5 - 1}{1.167 \times 10^4} = 3.428 \times 10^{-4} \, M$$

9.11.4 Marine Diatoms Gas Sensors

Recognizing the nanostructures and morphologies that nature has optimized during the long history of life on our planet helps in framing a completely different approach to engineering systems at the nanoscale. Diatoms are an important group of algae, and one of the common types of phytoplankton, found in the oceans, in freshwater, in soils and on damp surfaces. Stefano et al. (2005) noticed that the photoluminescence emission from silica skeleton of marine diatoms *Thalassiosira rotula* Meunier was strongly dependent on the surrounding environment. First, let us *see what electrophilic and nucleophilic substances are*. "Electrophilic" is the combination of two words (electro means electron and phile means loving). Therefore, electrophilic substances are electron lovers and are those substances that contain positive charges on them. Likewise, nucleophilic substances are nucleus lovers. They contain an atom or molecule with an excess of electrons and thus a negative charge.

In the work mentioned earlier, the researchers examined effect of diatom exposure to photoluminescence quenching gaseous substances: NO_2, acetone (C_3H_6O), and ethanol (C_2H_6O). All these substances are electrophiles so that they can attract some electrons from the silica skeleton of diatoms and quench its photoluminescence. The reverse effect was observed by exposing the diatoms to nucleophilic substances like xylene, C_8H_{10}; $C_6H_4(CH_3)_2$ or $C_6H_4C_2H_6$; and pyridine, C_5H_5N.

Thus, depending on their electronegativity and polarizing ability, some substances quench the luminescence, while others effectively enhance it. These phenomena allow reliable discrimination between different substances. This effect can be used for optical chemical sensing.

9.12 Discussion and Conclusions

Different types of gas nanosensors are tabulated in Table 9.1. These are grouped into three broad categories. First group of nanosensors is the chemiresistor class using metallic/MOX /CNT NPs. Other nanomaterials include polymer thin films, nanofibers, etc. The second class of nanosensors are the field-effect chemical sensors using CNTs, Si NWs, or ZnS/silica nanocables. This includes both gas and pH sensors in liquid media. The third type of nanosensors consists of ion sensors utilizing QD luminescence quenching. As demanded by the particular situation, the engineer chooses one from the various options available, e.g., if room-temperature operation is essential, then obviously MOXs working at high temperatures are unsuitable. Similar criteria are applied to make the correct choices.

TABLE 9.1

Diversity of Chemical Sensors

Sl. No.	Name of the Nanosensor	Sensed Quantity/ Applications	Nanomaterial/ Nanostructure Used	Sensing Principle
1.	Metallic nanoparticle-based gas sensor	Gas	Metallic nanoparticle	Conductivity change (chemiresistor)
2.	Metal oxide (MOX) gas sensor	Gas	Metal oxide nanoparticle	Conductivity change (chemiresistor)
3.	Carbon nanotube (CNT) gas sensor	Gas	CNT	Conductivity change (chemiresistor) / field-effect (FET)
4.	Porous silicon-based gas sensor	Gas	Nanoporous silicon	Conductivity change (chemiresistor)
5.	Thin organic polymer film–based gas sensor	Gas	Organic polymer	Conductivity change (chemiresistor)
6.	Electrospun polymer nanofibers as humidity sensor	Humidity	Polymer nanofiber	Conductivity change (chemiresistor)
7.	CNT-based ISFET for nano pH sensor	pH	CNT	Field effect
8.	NW nanosensor for pH detection	pH	Si NW	Field effect
9.	ZnS/silica nanocable FET pH sensor	pH	ZnS/silica nanocable	Field effect
10.	Bridging nanowire as vapor sensor	Vapor	Nanowire	
11.	Palladium functionalized Si NW H_2 sensor	H_2	Si NW	Field effect
12.	Polymer functionalized piezoelectric-FET humidity nanosensor	Humidity	Polymer	Effect of moisture-induced strain on piezoelectric gate
13.	Quantum dot ECL sensor	Metal ion	QD	Electrochemical luminescence
14.	BSA-activated CdTe QD nanosensor	Sb^{3+} ion	QD	Fluorescence quenching
15.	Functionalized CdSe/ZnS QD nanosensor	Hg(II) ion	QD	Luminescence quenching

Review Exercises

9.1 What is a chemical nanosensor? How does it differ from any macroscopic chemical sensor? Elaborate the statement "Detection of chemical processes on a single molecule scale is the ultimate goal of sensitive analytical assays."

9.2 Many chemical sensors operate through the variation of a surface parameter, like surface conductivity, with analyte concentration. Hence, the effective surface area of the device, i.e., the area actually interacting with the analyte, determines the sensitivity. How do nanomaterials provide an easy answer for increasing the surface area? Discuss the role played by different nanomaterials like NPs, nanotubes, nanowires, etc., in chemical sensors.

9.3 Highlight the importance of gas sensing. What are the gases that commonly need to be detected? What roles do nanomaterials play in fabricating improved gas-sensing devices?

9.4 Explain the working of a palladium NW sensor. How do palladium NWs enhance the performance of hydrogen sensors?

9.5 Why do MOX-based gas sensors need high temperatures for their normal functioning? What will happen if the required temperature is not provided?

9.6 What different types of hotplates are used for constructing metal oxide gas sensors? Present a comparison between the common types of miniaturized hotplates.

9.7 What are the attractive features of CNTs for gas-sensing applications? Explain the two approaches that are commonly followed for fabricating CNT-based gas sensors.

9.8 Are gas sensors based on raw CNTs selective for gas detection? How are CNTs functionalized for intended application?

9.9 Discuss in view of the following two remarks why the CNT-FET device presents serious impediments to commercialization: (i) Selective growth of metallic versus semiconducting CNTs is not possible with present technology. (ii) Further, if an *in situ* CVD process is used in the CNT-FET fabrication sequence, it is intricate to make a single SWCNT grow horizontally, bridging a given distance between the source and the drain so that one is forced to "pick and place" a semiconducting SWCNT from bulk samples.

9.10 DEP is a process through which neutral particles, such as CNTs, are translated through a suspending medium in a nonuniform electric field that is generated between a pair of electrodes. Name a sensor that is fabricated using this technique.

9.11 Describe a nano-FET constructed with palladium coating on MWCNTs for hydrogen sensing.

9.12 What is a microgravimetric sensor? How does CNT-based QCM sensor for ethanol work?

9.13 Describe a nanoporous silicon hydrogen sensor. Is it operated at room temperature? Why?

9.14 Explain the use of electrospun polymer nanofibers for humidity sensing. How does the modification of gold electrode with PDDA before nanofiber deposition help in improving the sensor performance?

9.15 Describe the use of a platform incorporating In_2O_3, SnO_2, and ZnO NWs, and SWCNTs together with micromachined hotplate for nano-electronic nose implementation.

9.16 How is a CNT-based ISFET used as a pH nanosensor? By what technique is the CNT aligned with contacting electrodes?

9.17 What surface-terminating functional groups are produced on a silicon NW by covalently linking APTES to it? Explain with structural formulae, the pH sensing mechanism of an APTES-modified silicon NW.

9.18 A higher Si–O asymmetric stretching bond density exists on the surface of the Si NWs sensing layer, than on the Si bulk materials in the same magnitude sensing window. The Si–O asymmetric stretching bonds can act as sites for hydrogen ion adsorption. Present arguments to convince that the pH sensitivity of silicon bulk materials is greatly improved by downsizing the sensors to the nanoscale.

9.19 How does a ZnS/SiO_2 core/shell nanocable FET work as a pH sensor? With the help of structural formulae, illustrate the operation of a biotin-functionalized nanocable FET for detecting streptavidin.

9.20 How does a bridging silicon NW sensor with Si electrodes avoid the difficulties of establishing contacts with metal electrodes interfacing with signal processing circuitry?

9.21 Explain the construction and operating mechanism of a Pd-functionalized Si NW sensor for hydrogen.

9.22 How does a polymer-functionalized zinc oxide FET sense moisture? Explain with diagrams the mechanisms responsible for generation of a humidity-dependent strain in the piezoelectric NB?

9.23 How is QD emission switched off in the antimony ion nanosensor in the presence of antimony? Illustrate your answer with diagrams.

9.24 How is mercury detected by analyte-induced changes in the photoluminescence of suitably modified QDs?

9.25 Explain the following statement: "NW sensors are highly sensitive because a significant fraction of the carriers in the NW can be depleted; therefore a small number of carriers induced by the analyte alters the conductance of NW appreciably."

References

Cattanach, K., R. D. Kulkarni, M. Kozlov, and S. K. Manohar. 2006. Flexible carbon nanotube sensors for nerve agent simulants. *Nanotechnology* 17: 4123–4128, doi: 10.1088/0957-4484/17/16/022

Chen, Z. H., J. S. Jie, L. B. Luo, H. Wang, C. S. Lee, and S. T. Lee. 2007. Applications of silicon nanowires functionalised with palladium nanoparticles in hydrogen sensors. *Nanotechnology* 18: 345502 (5 pp.), doi: 10.1088/0957-4484/18/34/345502

Cheng, L., X. Liu, J. Lei, and H. Ju. 2010. Low-potential electrochemiluminescent sensing based on surface unpassivation of CdTe quantum dots and competition of analyte cation to stabilizer. *Analytical Chemistry* 82: 3359–3364.

Cui, Y., Q. Wei, H. Park, and C. M. Lieber. 2001. Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species. *Science* 293(17): 1289–1292.

Dong, Z. and U. C. Wejinya. 2009. Design, fabrication and measurement of CNT based ISFET for NANO devices. *IEEE Nanotechnology Materials and Devices Conference*, June 2–5, 2009, Traverse City, MI, pp. 178–182.

Dong, Z., U. C. Wejinya, H. Yu, and I. H. Elhajj. 2009. Design, fabrication and testing of CNT based ISFET for NANO pH sensor application: A preliminary study. *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, Suntec Convention and Exhibition Center, Singapore, July 14–17, 2009, pp. 1556–1561.

Favier, F., E. C. Walter, M. P. Zach, T. Benter, and R. M. Penner. 2001. Hydrogen sensors and switches from electrodeposited palladium mesowire arrays. *Science* 293(5538): 2227–2231.

Fields, L. L., J. P. Zheng, Y. Cheng, and P. Xiong. 2006. Room temperature low-power hydrogen sensor based on a single tin dioxide nanobelt. *Applied Physics Letters* 88(26): Article ID 263102.

Francia, G. D., B. Alfano, and V. L. Ferrara. 2009. Conductometric gas nanosensors. *Journal of Sensors* 2009: Article ID 659275, 18 pp., doi: 10.1155/2009/659275

Ge, S., C. Zhang, Y. Zhu, J. Yu, and S. Zhang. 2010. BSA activated CdTe quantum dot nanosensor for antimony ion detection. *Analyst* 135: 111–115.

He, Jr. H., Y. Y. Zhang, J. Liu, D. Moore, G. Bao, and Z. L. Wang. 2007. ZnS/Silica nanocable field effect transistors as biological and chemical nanosensors. *Journal of Physical Chemistry C* 111(33): 12152–12156.

Hieu, N. V., L. T. B. Thuy, and N. D. Chien. 2008. Highly sensitive thin film NH_3 gas sensor operating at room temperature based on SnO_2/MWCNTs composite. *Sensors and Actuators B* 129: 888–895.

Hsu, J.-F., B.-R. Huang, C.-S. Huang, and H.-L. Chen. 2005. Silicon nanowires as pH sensor. *Japanese Journal of Applied Physics* 44(4B): 2626–2629.

Huang, X.-J. and Y.-K. Choi. 2007. Chemical sensors based on nanostructured materials. *Sensors and Actuators B* 122: 659–671.

Jun, J. H., J. Yun, K. Cho, I.-S. Hwang, J.-H. Lee, and S. Kim. 2009. Necked ZnO nanoparticle-based NO_2 sensors with high and fast response. *Sensors and Actuators B: Chemical* 140: 412–417.

Kamins, T. I., S. Sharma, A. A. Yasseri, Z. Li, and J. Straznicky. 2006. Metal-catalysed, bridging nanowires as vapour sensors and concept for their use in a sensor system. *Nanotechnology* 17: S291–S297, doi: 10.1088/0957-4484/17/11/S11

Lao, C. S., Q. Kuang, Z. L. Wang, M.-C. Park, and Y. Deng. 2007. Polymer functionalized piezoelectric-FET as humidity/chemical nanosensors. *Applied Physics Letters* 90: 262107-1–262107-3.

Law, M., H. Kind, B. Messer, F. Kim, and P. Yang. 2002. Photochemical sensing of NO_2 with SnO_2 nanoribbon nanosensors at room temperature. *Angewandte Chemie International Edition* 41(13): 2405–2408.

Li, P., Y. Li, B. Ying, and M. Yang. 2009. Electrospun nanofibers of polymer composite as a promising humidity sensitive material. *Sensors and Actuators B: Chemical* 141(2): 390–395.

Li, J., Y. Lu, and M. Meyyappan. 2006. Nano chemical sensors with polymer-coated carbon nanotubes. *IEEE Sensors Journal* 6(5): 1047–1051.

Li, J., Y. Lu, Q. Ye, M. Cinke, J. Han, and M. Meyyappan. 2003. Carbon nanotube sensors for gas and organic vapor detection. *Nano Letters* 3(7): 929–933.

Li, H., Y. Zhang, X. Wang, and Z. Gao. 2008. A luminescent nanosensor for Hg(II) based on functionalized CdSe/ZnS quantum dots. *Microchim Acta* 160: 119–123, doi: 10.1007/s00604-007-0816-x

Lu, Y., J. Li, J. Han, H.-T. Ng, C. Binder, C. Partridge, and M. Meyyappan. 2004. Room temperature methane detection using palladium loaded single-walled carbon nanotube sensors. *Chemical Physics Letters* 391(4–6): 344–348.

Lucci, M., A. Reale, A. Di Carlo, S. Orlanducci, E. Tamburri, M. L. Terranova, I. Davoli, C. Di Natale, A. D'Amico, and R. Paolesse. 2006. Optimization of a NOx gas sensor based on single walled carbon nanotubes. *Sensors and Actuators B: Chemical* 118(1–2): 226–231.

Lucci, M., P. Regoliosi, A. Reale, A. Di Carlo, S. Orlanducci, E. Tamburri, M. L. Terranova, P. Lugli, C. Di Natale, A. D'Amico, and R. Paolesse. 2005. Gas sensing using single wall carbon nanotubes ordered with dielectrophoresis. *Sensors and Actuators B* 111–112: 181–186.

Luongo, K., A. Sine, and S. Bhansali. 2005. Development of a highly sensitive porous Si based hydrogen sensor using Pd nano-structures. *Sensors and Actuators B* 111–112: 125–139.

Ma, Y., W. L. Wang, K. J. Liao, and C. Y. Kong. 2002. Study on sensitivity of nano-grain ZnO gas sensors. *Journal of Wide Bandgap Materials* 10(2): 113–120.

Massera, E., I. Nasti, L. Quercia, I. Rea, and G. Di Francia. 2004. Improvement of stability and recovery time in porous silicon-based NO_2 sensor. *Sensors and Actuators B* 102(2): 195–197.

Nohria, R., R. K. Khillan, Y. Su, R. Dikshit, Y. Lvov, and K. Varahramyan. 2006. Humidity sensors based on polyaniline film deposited using layer-by-layer nano-assembly. *Sensors and Actuators B* 114: 218–222.

Padigi, S. K., R. K. K. Reddy, and S. Prasad. 2007. Carbon nanotube based aliphatic hydrocarbon sensor. *Biosensors and Bioelectronics* 22(6): 829–837.

Penza, M., G. Cassano, P. Aversa, F. Antolini, A. Cusano, A. Cutolo, M. Giordano, and L. Nicolais. 2004. Alcohol detection using carbon nanotubes acoustic and optical sensors. *Applied Physics Letters* 85(12): 2379–2381.

Plessis, M. du. 2008. Nanoporous silicon explosive devices. *Materials Science and Engineering B* 147(2–3): 226–229.

Qi, P., O. Vermesh, M. Grecu, A. Javey, Q. Wang, H. Dai, S. Peng, and K. J. Cho. 2003. Toward large arrays of multiplex functionalized carbon nanotube sensors for highly sensitive and selective molecular detection. *Nano Letters* 3(3): 347–351.

Shen, Y., T. Yamazaki, Z. Liu, D. Meng, T. Kikuta, N. Nakatani, M. Saito, and M. Mori. 2009. Microstructure and H_2 gas sensing properties of undoped and Pd-doped SnO_2 nanowires. *Sensors and Actuators B* 135(2): 524–529.

Stefano, L. D., I. Rendina, M. D. Stefano, A. Bismuto, and P. Maddalena. 2005. Marine diatoms as optical chemical sensors. *Applied Physics Letters* 87: 233902-1–233902-2.

Suehiro, J., G. Zhou, H. Imakiire, W. Ding, and M. Hara. 2005. Controlled fabrication of carbon nanotube NO_2 gas sensor using dielectrophoretic impedance measurement. *Sensors and Actuators B* 108(1–2): 398–403.

Walter, E. C., F. Favier, and R. M. Penner. 2002a. Palladium mesowire arrays for fast hydrogen sensors and hydrogen actuated switches. *Analytical Chemistry* 74(7): 1546–1553.

Walter, E. C., K. Ng, M. P. Zach, R. M. Penner, and F. Favier. 2002b. Electronic devices from electrodeposited metal nanowires. *Microelectronic Engineering* 61–62: 555–561.

Wong, Y. M., W. P. Kang, J. L. Davidson, A. Wisitsora-at, and K. L. Soh. 2003. A novel microelectronic gas sensor utilizing carbon nanotubes for hydrogen gas detection. *Sensors Actuators B: Chemical* 93(1): 327–332.

Yamazoe, N. 2005. Toward innovations of gas sensor technology. *Sensors and Actuators B* 108: 2–14.

Yamazoe, N. and K. Shimanoe. 2008. Roles of shape and size of component crystals in semiconductor gas sensors: I. Response to oxygen. *Journal of the Electrochemical Society* 155(4): J85–J92.

Yamazoe, N. and K. Shimanoe. 2009. Receptor function and response of semiconductor gas sensor. *Journal of Sensors* 2009: Article ID 875704, 21 pages, doi: 10.1155/2009/875704

Zhang, B., R. Fu, M. Zhang, X. Dong, B. L. Wang, and C. U. Pittman. 2006a. Gas sensitive vapor grown carbon nanofiber/polystyrene sensors. *Materials Science Bulletin* 41: 553–562.

Zhang, T., M. B. Nix, B.-Y. Yoo, M. A. Deshusses, and N. V. Myung. 2006b. Electrochemically functionalized single-walled carbon nanotube gas sensor. *Electroanalysis* 18(12): 1153–1158.

Zhang, W.-D. and W.-H. Zhang. 2009. Carbon nanotubes as active components for gas sensors. *Journal of Sensors* 2009: Article ID 160698, 16 pp., doi: 10.1155/2009/160698

10

Future Trends of Nanosensors

10.1 Introduction

The progression from macro- to nanosensors has been a very long and interesting evolution (Khanna 2008, 2009). Several nanomaterials, notably metallic nanoparticles (NPs), carbon nanotubes (CNTs), and quantum dots (QDs), together with enabling nanotechnologies, duly supported by microtechnologies have been exploited to realize nanosensors.

10.1.1 Interfacing Nanosensors with Human Beings

The core idea is to understand that a nanosensor is not necessarily a device merely reduced in size to a few nanometers, but a device that makes use of the unique properties of nanomaterials and NPs to detect and measure new types of events and entities in the nanoscale. Nanotechnology received a boost from the renewed interest in colloidal science combined with the introduction of a new generation of analytical tools such as scanning tunneling and atomic force microscope (AFM). Deliberate manipulation of nanostructures by electron-beam lithography and molecular beam epitaxy enabled the observation of novel nanoscale phenomena.

An optical fiber pierces the cell wall nondestructively to peep into what is happening inside the cell. The active portion of the sensor penetrating the cell wall has to be of nano dimensions to avoid damage to the cell, but if its connection to the outside world is of that scale, no one will be able to see and operate it. The interface between human beings and nanosensors is of utmost importance. Ultimately, whether it is through an instrument or some improvisation, the nanosensor has to be connected to the macro world.

Of serious concern are the compatibility issues at the borderline of human beings and nanosensors. After all, the nanosensor has to work in this macroworld and human beings are neither microscopic nor nanoscopic creatures although our structural units are of those scales. *Could we manipulate NPs if necessary tools were not developed?* No, we can neither see nor work with nanoobjects without aiding tools. All nanoscale sensors have to be operated by our macroscopic-level capabilities using convenient aids. Man is a tool-using animal.

10.1.2 Three Main Types of Nanosensors

Nanosensors belong to one of the three types: physical, chemical, and biological. Physical nanosensors are used to measure magnitudes such as mass, pressure, force, or displacement. Chemical nanosensors are employed to measure parameters such as the concentration of a given gas, the presence of a specific type of molecules, or the molecular composition of a substance. Biological nanosensors are used to monitor biomolecular processes such as antibody/antigen interactions, DNA interactions, enzymatic interactions, or cellular communication processes.

10.1.3 Using the Response Properties of the Same Nanomaterial in Different Types of Nanosensors

Can the same nanomaterial act as the central element of different types of nanosensors? Yes, the same nanomaterial, for example, CNT, can be used as the key element of physical, chemical, and biological nanosensors (Akyildiz and Jornet 2010); CNT-based sensors have been described in almost all chapters. A CNT can be used to build a field-effect transistor (FET) whose on/off threshold changes on a local deformation of the tube serving as a force sensor (mechanical sensor). CNTs are used to sense chemical compounds. When used as a gas sensor (chemical sensor), the presence of a specific type of gas molecules changes the threshold voltage of the transistor. The CNT biological sensors work in a similar way to chemical nanosensors, but in this case, the change in the electronic properties of, for example, a CNT-based FET transistor is induced by biomolecules, by either a protein or any other chemical composite that binds itself to the functionalized nanotube, a specific antigen that binds itself to an antibody glued to the nanotube or a single-stranded DNA chain that binds itself to another DNA chain which has been attached to the nanotube. But the most important aspect of the investigation of a variety of sensors is three S's, that is, sensitivity, selectivity, and stability.

10.1.4 Nanosensor Science, Engineering, and Technology: Three Interrelated Disciplines

As we know, science comprises the rules of the game, engineering is actually playing the game, and technology contains the tips that help us win the game. In this background, there are three important disciplines in reference to nanosensors: nanosensor science, nanosensor engineering, and nanosensor technology. The first discipline involves appreciating the nanolevel phenomena and laws of nature that can be applied for designing various nanosensors. This is the field of nanosensor scientists. The second discipline deals with application of these laws by nanosensor engineers to construct nanosensors according to the desired application. The third discipline

constitutes the skills acquired for successful fabrication of vast variety of nanosensors described in preceding pages of this book. This is the intellectual property of nanosensor technologists who have been granted patents for their inventions.

10.1.5 Scope of the Chapter

This chapter will provide a summary of the present status of main types of nanosensors: physical, chemical, and biological, which have been described in this book, point out their current capabilities and limitations, and highlight the research problems that need to be addressed in this area.

10.2 Scanning Tunneling Microscope

In a scanning tunneling microscope (STM) (Giessibl 2003), a pointed tip is brought near an electrically conducting surface biased at a voltage V_t. When the separation between the tip and the surface is very small, typically a few atomic diameters, electrons are transported by tunneling and a current I_t flows between them. The strong distance dependence of tunneling current lays the foundation of the atomic resolution capability of the STM. However, the two objects involved in tunneling must be conducting or at least semiconducting but not insulating, imposing a restriction on the nature of specimen that can be examined.

10.3 Atomic Force Microscope

The main component of an AFM (Giessibl 2003) and its major instrumental difference from an STM is the spring used for sensing the force between tip and sample. *What requirements should be met by the force sensor for sensing normal tip–sample forces?* It must be rigid in two axes and relatively soft in the third axis. *Which device realizes this property?* It is fulfilled by a cantilever beam, and, therefore, the cantilever geometry is typically used for force detection. Access to the nanoworld is gained by means of a local probe operating with *surface contact* or *in a near-field mode* at a few angstroms sample–tip distance only. The first AFMs were mostly operated in the *static contact mode*, and for this mode, the restriction on the stiffness of the cantilever is that it should be less than the interatomic spring constants of atoms in a solid, which is $k \leq 10\,\mathrm{N\,m^{-1}}$. In *dynamic atomic force microscopy*, k values higher than hundreds of N m^{-1} assist in reducing noise

and increasing operational stability. The Q factor depends on the damping mechanisms present in the cantilever. For micromachined cantilevers operated in air, Q is principally limited by viscous drag of air and characteristically amounts to a few hundred. In vacuum, internal and surface effects in the cantilever material are responsible for damping and Q values of hundreds of thousands are commonplace. Mammoth progress has been made in fabricating attonewton-sensitive cantilevers from both silicon nitride and silicon.

10.4 Mechanical Nanosensors

The ability to measure and quantify the motion of an object is a basic sense required in advanced control systems. Mechanical nanosensors are of fundamental scientific interest. They have been employed in applications including ultrasensitive displacement, acceleration, force, and resonant mass sensing of chemical and biological species.

How was tunneling phenomenon applied in developing mechanical or other nanosensors? The tunneling phenomenon was used in various nanosensors for displacement, acceleration, and magnetic field. The miniature high-sensitivity accelerometers have been developed using a high-resolution displacement transducer based on electron tunneling. In a MEMS (microelectromechanical systems) technology-based tunneling accelerometer, the tunneling current between the tip and counter electrode is maintained constant through a feedback system. The electrical force used to hold the current constant is measured and converted into an acceleration measurement. Magnetic fields are measured through an indirect application of tunneling.

How have MEMS/NEMS technologies contributed toward nanosensors? With microelectronics technology pushing deep into the sub-micrometer size regime, a concerted effort surfaced to realize even smaller electromechanical systems: nanoelectromechanical systems (NEMS). NEMS convert electrical current into mechanical motion on a nanoscale and vice versa (Ekinci 2005). They can be viewed as the descendants of MEMS, which operate at a micron scale and which are found in commercial applications. Improved performance is expected from NEMS devices due to their small sizes and higher eigen frequencies.

What are the important NEMS devices used as nanosensors? NEMS devices include cantilevered and doubly clamped beam resonators operated in their fundamental flexural resonant modes, much like simple tuning forks. In this size regime, NEMS offer access to GHz frequencies, with quality factors in excess of 10^4. In general, the smaller a device, the more susceptible are its physical properties to perturbation by external influences. This enhanced

sensitivity of NEMS is opening a variety of unprecedented opportunities for applications such as mass spectrometry. The vibrational frequency of a NEMS resonator is an exquisitely sensitive function of its total mass. Small variations in mass, such as resulting from adsorbed addenda, can measurably alter its resonant frequency. Theoretical calculations for physically realizable devices indicate that NEMS mass sensitivity below a single Dalton (1 Da = AMU) is achievable.

What constructional materials have been used in NEMS resonators? NEMS resonating structures currently used are based on single-crystal Si fabricated from silicon-on-insulator substrates, because the mechanical and electrical properties of these systems are well known and processing techniques are well developed. The combination of electron beam lithography and Si micromachining techniques has made it possible to fabricate submicron mechanical structures from single crystal substrates. There have been efforts to utilize other materials with particularly high Young's modulus, such as aluminum nitride, silicon carbide, and ultra nanocrystalline diamond.

What are the chief resonant mass nanosensors? Resonant mass nanosensors are already employed in many assorted fields of science and technology. These devices operate by providing a frequency shift that is directly proportional to the inertial mass of the analyte molecules deposited upon them. Among the most sensitive realizations are those based on the acoustic vibratory modes of crystals, thin films, and micron-sized cantilevers. Quartz crystal microbalances (QCM) utilize the characteristic of frequency variation with mass loading of the resonator. A QCM sensor can detect up to 10^{-13} g, and is used to probe nanometer thick film loads. QCM is capable of measuring mass changes as small as a fraction of a monolayer or single layer of atoms.

How have MEMS/NEMS-based balances been used? MEMS/NEMS with various designs and approaches have been used as balances to detect the mass of the species attached to the NEMS structure with mass sensitivities ranging from picograms (10^{-12} g) to zeptograms (10^{-21} g). Self-sensing cantilevers with metallic piezoresistors, having dimensions approaching the mean free path at atmospheric pressure, maintain high resonance quality factors in ambient conditions. This enables measurements in air at room temperature, with unmatched mass resolution less than 1 ag (10^{-18} g).

How has single-electron transistor been applied in displacement sensing? Single-electron transistor (SET) has gained renown for its exceptional sensitivity to charge. The most outstanding property of SET is the possibility to switch the device from the insulating to the conducting state by adding only one electron to the gate electrode. The elegant charge sensitivity of SET at cryogenic temperatures is exploited to measure displacement by capacitively coupling it to the mechanical resonator (Knobel and Cleland 2003). The detection of the piezoelectrically induced charge in a mechanical resonator with a SET has been shown to be a prime candidate for nearly quantum-limited displacement sensing. If instead, the resonator is fabricated from a piezoelectric

material, such as GaAs, AlGaAs, or AlN, and the SET is configured to sense the piezoelectric voltage developed when the beam flexes, then a considerably higher displacement sensitivity can be achieved.

What are the applications of CNTs for determining mechanical parameters? CNTs act as sensing materials in force, pressure, flow, mass, position, stress, and strain sensors. (i) A tunable doubly clamped CNT electromechanical oscillator executes guitar-string-like oscillation modes. Its resonance frequency can be widely tuned and the devices can be used to *transduce very small forces.* (ii) *Atomic scale mass sensing* has been accomplished using doubly clamped suspended CNT nanomechanical resonators, in which their SET properties allows self-detection of the nanotube vibration. The detection of shifts in the resonance frequency of the nanotubes has been applied to sense and determine the inertial mass of atoms as well as the mass of the nanotube. (iii) The band structure of a CNT is drastically altered by *mechanical strain* leading to a strain nanosensor. (iv) The rate of *fluid flow* around a CNT sensor affects the heat transfer between the CNT sensor and the environment. In its turn, this heat transfer affects the power required to heat the CNT sensor to a specific temperature; hence, CNT sensor carefully biased for self-heating is used for fluid-flow rate sensors (Chow et al. 2010). (v) CNT transistors, incorporated into microfluidic channels, locally sense the change in electrostatic potential induced by the flow of an ionic solution (Bourlon et al. 2007).

Example 10.1

A cantilever having a spring constant of 0.01 N m^{-1}, resonance frequency of 1 MHz is vibrating with an amplitude of 3×10^{-6} m at 27°C. The quality factor is 10^4 and the bandwidth is 1 Hz. Calculate: (a) the minimum detectable frequency change, (b) the minimum detectable force, and (c) the mass resolution achievable with this structure.

  (a) The frequency shift is given by

$$\delta\omega = \left(\frac{1}{A}\right)\left(\frac{\omega_0 k_B TB}{kQ}\right)^{0.5} \tag{10.1}$$

  where
  A is the amplitude of oscillation
  ω_0 is the resonance frequency
  k_B is Boltzmann constant
  T is temperature
  B is bandwidth
  k is spring constant of cantilever
  Q is the quality factor of cantilever

Here, $A = 3 \times 10^{-6}$ m, $\omega_0 = 2\pi \times 10^6$ Hz $= 6.28 \times 10^6$ Hz, $k_B = 1.381 \times 10^{-23}$ J K^{-1}, $T = 27 + 273 = 300$ K, $B = 1$ Hz, $k = 0.01$ N m^{-1}, $Q = 10^4$.

$$\delta\omega = \left(\frac{1}{3 \times 10^{-6}}\right)\left(\frac{6.28 \times 10^6 \times 1.381 \times 10^{-23} \times 300 \times 1}{0.01 \times 10^4}\right)^{0.5}$$

$$= 5.3767 \times 10^{-3} \text{ s}^{-1} \tag{10.2}$$

Minimum detectable angular frequency shift is 5.3767×10^{-3} Hz.

(b) The minimum force that can be detected by a cantilever is limited by thermomechanical fluctuations to

$$F_{\text{minimum}} = \left(\frac{2kk_B TB}{\pi Q f_0}\right)^{0.5} \tag{10.3}$$

where
k_B is Boltzmann's constant $= 1.381 \times 10^{-23}$ J K^{-1}
T is temperature $= 300$ K
B is the detection bandwidth $= 1$ Hz
$k = 0.01$ N m^{-1}
$f_0 = 10^6$ Hz
$Q = 10^4$ are cantilever spring constant, resonance frequency, and quality factor, respectively

Putting the values in Equation 10.3, we get

$$F_{\text{minimum}} = \left(\frac{2kk_B TB}{\pi Q f_0}\right)^{0.5} = \left(\frac{2 \times 0.01 \times 1.381 \times 10^{-23} \times 300 \times 1}{3.14 \times 10^4 \times 10^6}\right)^{0.5}$$

$$= 5.137 \times 10^{-17} = 51.37 \text{ aN} \tag{10.4}$$

(c) The expected mass resolution is

$$\delta m = k\left(\frac{1}{\omega_1^2} - \frac{1}{\omega_2^2}\right) \tag{10.5}$$

where
ω_1 is the initial resonance frequency of cantilever $= 6.28 \times 10^6$ Hz
ω_2 is the final resonance frequency of cantilever $= 6.28 \times 10^6 + 6.28 \times 5.3767 \times 10^{-3} = 6.28 \times 1.0000000053767 \times 10^6$ Hz

Therefore,

$$\delta m = k\left(\frac{1}{\omega_1^2} - \frac{1}{\omega_2^2}\right) = \frac{0.01}{(6.28)^2}\left\{\frac{1}{(1\times10^6)^2} - \frac{1}{(1.0000000053767\times10^6)^2}\right\}$$

$$= \frac{0.01}{(6.28)^2}\left(1\times10^{-12} - 9.999999892466\times10^{-13}\right)$$

$$= 2.54\times10^{-4}\times1.07534\times10^{-20}\ \text{kg}$$

$$= 2.731\times10^{-24}\ \text{kg} \tag{10.6}$$

Example 10.2

An $Al/Al_2O_3/Al$ SET is fabricated. For optimal performance of the SET, the total island capacitance must be minimized, that is, the area of the tunnel junctions must be as small as possible. Taking the total central island capacitance as 200×10^{-18} F, what is the operating temperature of the SET device? ($k_B = 1.381\times10^{-23}$ J K^{-1}).

For SET operation based on Coulomb blockade, the relation,

$$k_B T \ll \frac{q^2}{2C} \tag{10.7}$$

must be satisfied, that is,

$$T \ll \frac{q^2}{2k_B C} = \frac{(1.6\times10^{-19})^2}{2\times1.381\times10^{-23}\times200\times10^{-18}} = 4.634\ \text{K} \tag{10.8}$$

Therefore, for ensuring reliable operation, the temperature must be <1 K.

Example 10.3

Find the resonance frequency of a doubly clamped beam resonator made of gallium arsenide having the following dimensions: length = 8 μm, width = 600 nm, and thickness = 259 nm, given that the Young's modulus (E_{GaAs}) of gallium arsenide is 85.5 GPa and its density (ρ_{GaAs}) is 5317 kg m^{-3}. Will the resonance frequency of a silicon beam be lower or higher than that for gallium arsenide, if $E_{Si} = 185$ GPa and $\rho_{Si} = 2328$ kg m^{-3}.

What is the ratio of resonance frequencies for the two materials? The fundamental resonance frequency, f, of a doubly clamped beam of length, L, and thickness, t, varies linearly with the geometric factor t/L^2 according to the simple relation

$$f = 1.03\sqrt{\frac{E}{\rho}}\left(\frac{t}{L^2}\right) \tag{10.9}$$

where
  E is the Young's modulus
  ρ is the mass density

Note that beam width (w) does not enter the calculations.
  $1\,\mathrm{Pa} = 1\,\mathrm{N\,m^{-2}}$. Hence, $E_{\mathrm{GaAs}} = 85.5\,\mathrm{GPa} = 85.5 \times 10^9\,\mathrm{N\,m^{-2}} = 8.55 \times 10^{10}\,\mathrm{N\,m^{-2}}$. $E_{\mathrm{Si}} = 185\,\mathrm{GPa} = 185 \times 10^9\,\mathrm{N\;m^{-2}} = 1.85 \times 10^{11}\,\mathrm{N\;m^{-2}}$. From Equation 10.9, the resonance frequency of a silicon beam is

$$f_{\mathrm{Si}} = 1.03\sqrt{\frac{E_{\mathrm{Si}}}{\rho_{\mathrm{Si}}}}\left(\frac{t}{L^2}\right) = 1.03\sqrt{\frac{1.85 \times 10^{11}}{2328}}\left\{\frac{\left(259 \times 10^{-9}\right)}{\left(8 \times 10^{-6}\right)^2}\right\}$$

$$= 1.03 \times 8.9144 \times 10^3 \times 4.047 \times 10^3 = 3.716 \times 10^7 = 3.72 \times 10^7\ \mathbf{Hz} \quad (10.10)$$

while that of a GaAs beam is

$$f_{\mathrm{Si}} = 1.03\sqrt{\frac{E_{\mathrm{GaAs}}}{\rho_{\mathrm{GaAs}}}}\left(\frac{t}{L^2}\right) = 1.03\sqrt{\frac{8.55 \times 10^{10}}{5317}}\left\{\frac{\left(259 \times 10^{-9}\right)}{\left(8 \times 10^{-6}\right)^2}\right\}$$

$$= 1.03 \times 4.01 \times 10^3 \times 4.047 \times 10^3 = 1.672 \times 10^7\ \mathrm{Hz} \quad (10.11)$$

Silicon will give a higher resonance frequency than gallium arsenide because of its larger (Young's modulus/density) ratio. The ratio of resonance frequencies for the two materials is given by

$$\frac{f_{\mathrm{Si}}}{f_{\mathrm{GaAs}}} = \frac{1.03\left(t/L^2\right)\sqrt{E_{\mathrm{Si}}/\rho_{\mathrm{Si}}}}{1.03\left(t/L^2\right)\sqrt{E_{\mathrm{GaAs}}/\rho_{\mathrm{GaAs}}}} = \frac{\sqrt{E_{\mathrm{Si}}/\rho_{\mathrm{Si}}}}{\sqrt{E_{\mathrm{GaAs}}/\rho_{\mathrm{GaAs}}}} = \frac{8.9144 \times 10^3}{\left(4.01 \times 10^3\right)} = 2.223$$

$$(10.12)$$

which is the same as given by the ratio of Equations 10.10 and 10.11, that is, $3.716 \times 10^7/(1.672 \times 10^7) = 2.222$.

10.5 Thermal Nanosensors

Temperature nanosensor was made by electron-beam chemical vapor deposition (EB-CVD) using Pt source gas material. W thermal nanosensor with 3D nanostructure was made by FIB-CVD using $W(CO)_6$ source gas material (Ozasa et al. 2004).

How have CNTs been exploited as temperature sensors? (i) CNTs grown on nickel film deposited on float glass substrate serve as low-temperature sensors. (ii) Laterally grown CNT between two electrodes is used as a nano-temperature sensor. *Have carbon nanowires (NWs) been used as temperature sensors?* Yes, temperature sensors have been developed using an array of

carbon NWs written by Ga^+ focused ion beam. The NW technology obliter-ates the difficulty in selective growth or placement of CNT. *How is a CNT-based infrared sensor constructed?* The infrared photoresponse in the electrical conductivity of single-walled carbon nanotubes (SWCNTs) is spectacularly enhanced by embedding SWCNTs in an electrically and thermally insulat-ing polymer matrix, enabling the construction of an infrared sensor.

What is the status of silicon nanowires (Si NWs) as thermal sensors? Silicon NWs are employed as nano-temperature sensors in two configurations, that is, as (i) resistance temperature detector (RTD) and (ii) diode temperature detector (DTD) types. Si NWs reap the benefits of nanoelectronics technol-ogy with regard to geometrical patterning by nanolithographic techniques with attendant precise location advantage.

How are NPs utilized as temperature sensors? (i) A ratiometric fluorescent NP has been reported for temperature sensing using an alkoxysilanized dye as a reference. Er^{3+}-doped $BaTiO_3$ nanocrystals are suitable for use as FIR-based temperature nanosensors. (ii) Gd_2O_3: Er^{3+}/Yb^{3+} nanocrystalline phosphor has been used as a temperature sensor.

What are the promises of bolometers with reference to superconducting transition edge nanosensor? Since the original invention by Samuel P. Langley in 1878, bolometers have gone a long way of improving the sensitivity and expanding the frequency range, from x-rays and optical/UV radiation to the sub-millime-ter waves. The latter range contains approximately half the total luminosity of the universe and 98% of all the photons emitted since the Big Bang started the universe. Because the performance of ground-based THz telescopes is harshly limited by a strong absorption of THz radiation in the terrestrial atmosphere, the development of space-based THz telescopes will be crucial for better understanding of the universe evolution. Although new detector concepts are coming into play currently, bolometers still have a great potential for achieving the most challenging goals. The notion of improving the sensitivity of bolom-eters by employing the hot-electron effects at ultralow temperatures has been enthusiastically pursued (Wei et al. 2008). Hot-electron direct detector (HEDD) element is a superconducting transition edge nanosensor made from thin Ti film with superconducting transition temperature $T_C \sim 0.2$–$0.4\,K$. The current leads to the nanosensor are fabricated from Nb films with $T_C \sim 8.5\,K$; a large superconducting gap in Nb blocks outdiffusion of "hot electrons" to the cur-rent leads. The nanostructure is fabricated on a silicon substrate using electron-beam lithography and electron-gun deposition of Ti and Nb. The detector has low NEP $= 3 \times 10^{-19}\,W\,Hz^{-1/2}$ at $0.3\,K$, and is capable of reckoning THz photons.

10.6 Optical Nanosensors

What is the main refractive index change strategy used in optical nanosensors? The LSPR nanosensors induce the small local refractive index changes at the

surface of metallic NPs and function by transducing small changes in the refractive index near the metallic surface into a measurable wavelength shift response (Barbillon et al. 2007). *What are the prospects of SERS-based nanosensing?* Compared with the refractive index–based detection schemes, SERS is a vibrational spectroscopic method that yields unique vibrational signatures for small molecule analytes, as well as quantitative information (Kneipp et al. 2010). The SERS signal transduction mechanism has many characteristics that can be utilized in biosensing applications: sensitivity, selectivity, low laser power, and no interference from water molecules. Examples are trace analysis of DNA, bacteria, glucose, living cells, posttranslational modification of proteins, enzyme, chemical warfare agents, and CNTs. A miniaturized, inexpensive, and portable Raman instrument makes the technique practical for trace analysis in clinics, field, and urban settings.

Nanobiotechnology, a new word, describes the use of nanotechnology in biological systems. *How does the colloidal gold nanobiosensor work?* A colorimetric gold NP sensor for probing biomolecular interactions works on the change in the absorbance spectrum of a self-assembled monolayer of colloidal gold on glass, as a function of biomolecular binding to the surface of the immobilized colloids.

What are the capabilities of optical fiber nanobiosensors? Fiber-optic nanobiosensors are an inimitable class of biosensors that enable analytical measurements in individual living cells and the probing of individual chemical species in specific locations within a cell. Fiber-optic nanobiosensors consisting of antibodies, as biorecognition molecules, coupled to an optical transducer element, have been developed and used to detect biochemical targets, benzopyrene tetrol (BPT), and benzo[a]pyrene (BaP), inside single cells.

How are optical fibers utilized for optochemical sensing applications? Standard silica optical fibers (SOFs) coated by SWCNTs Langmuir–Blodgett multilayers are used as highly VOC sensors based on light reflectometry, that is, through film reflectance-induced changes. Metal oxides (MOX) are another interesting class of materials widely exploited in gas sensing applications. The term "reflectometric configuration" is generally used to refer to the standard extrinsic Fabry–Perot (SEFP) configuration, which relies on reflectance measurements from a low finesse and extrinsic Fabry–Perot interferometer, realized by depositing a flat thin sensitive layer onto the distal end of properly cut optical fibers. Nano-coated long period fiber gratings, modified Fabry–Perot interferometers involving near field effect and photonic bandgap modification in hollow-core optical fibers, have been proposed, respectively, to provide the best sensing performance.

How does the fluorescence-based pH nanosensor operate? A ratiometric pH nanosensor, contains a pH-sensitive fluorescent dye as well as a pH-insensitive reference dye, embedded in a polymer matrix. A charged fluorescent molecule, for example, a reporter dye, is trapped and concentrated in solution at the nanopipette tip. The dye is excited by focusing a laser beam using far-field optics at the 100 nm inner diameter pipette tip, producing a local nanosensor

where the fluorescence is dependent on analyte concentration in the bath. SNARF-1-dextran, a negatively charged ratiometric pH-sensitive fluorophore, is used to demonstrate the concept of the nanopipette as a pH sensor.

What are the vital attributes of PEBBLES? PEBBLEs (photonic explorers for bioanalysis with biologically localized embedding) are nanometer-sized spherical optical sensing devices made from biologically inert polymers and are capable of *in vitro* measurements (Sumner et al. 2006). These sensors are fabricated in a microemulsion and consist of fluorescent indicators entrapped in a polyacrylamide matrix. The sensing mechanism involves the permeation of analytes into the NP matrix and their selective interaction with sensing components, resulting in signal changes. Fluorescent indicator and reference dyes are loaded into the polymeric NP core or its surface-coated layer and there is no interaction between the dyes. Such NP sensors have been developed for H^+, Ca^{2+}, Mg^{2+} ions, OH radical, small molecules like O_2, etc.

What is electrochemiluminescent (ECL) nanosensing approach? In electrochemiluminescent nanosensor, ECL light is initiated by the gold nanoring electrode in the presence of a co-reactant biospecies, NADH (Chovin et al. 2006).

How does the UV nanosensor work? Naturally self-assembled crossed ZnO nanorods exhibit a response of ~15 mA W^{-1} for UV light (361 nm) under 1 V bias (Chai et al. 2009).

What is the impact of QD nanosensors in the sensing arena? Compared to conventional organic dye molecules, fluorescent semiconductor nanocrystals (QDs) have several promising advantages. A few examples are given as follows: (i) Quantitative maltose sensing by QD-FRET has shown the *modus operandi* by which QDs might play a role in enzyme assays. (ii) Several studies have exploited QD-FRET for imaging activity of proteases. (iii) FRET-based QD bioprobes are individually able to detect the actions of protease, deoxyribonuclease, DNA polymerase, or changes in pH. (iv) In addition, two such QD-mounted biosensors were excited at a single wavelength and shown to operate simultaneously and independently of each other in the same sample solution. Thus, multiplex detection of the action of a protease, trypsin, was possible in the presence of deoxyribonuclease.

Example 10.4

Surface plasmon resonance measures the change in the index refraction of the fluid medium near the sensor surface. For incident light of wavelength 635 nm, what is the difference in wavelength of emergent beams from two protein binding layers of refractive indices 1.5 and 1.515?

If the velocity of light in a medium is v and the same in vacuum is c, the relationship between the wavelength $\lambda = v/f$ in a medium and the wavelength $\lambda_0 = c/f$ in vacuum is

$$\frac{\lambda}{\lambda_0} = \frac{v/f}{c/f} = \frac{v}{c} \qquad (10.13)$$

The frequencies cancel because frequency does not change as light moves from one medium to another. Because refractive index n is defined as the ratio of the velocity of light in a vacuum to that in a medium, the above equation reduces to

$$\frac{\lambda}{\lambda_0} = \frac{1}{n} \qquad (10.14)$$

Hence,

$$\lambda = \frac{\lambda_0}{n} \qquad (10.15)$$

The refractive index n in this equation measured with respect to vacuum is called the *absolute refractive index*. In practice, air makes little difference to the refraction of light with an absolute refractive index of 1.0008, so the value of the absolute refractive index can be used assuming the incident light is in air. So, Equation 10.15 can be written for the first protein binding layer as

$$\lambda_{\text{protein1}} = \frac{\lambda_{\text{air}}}{n_{\text{protein1}}} = \frac{635}{1.5} = 423.33 \text{ nm} \qquad (10.16)$$

For the second protein layer, Equation 10.15 is rewritten as

$$\lambda_{\text{protein2}} = \frac{\lambda_{\text{air}}}{n_{\text{protein2}}} = \frac{635}{1.515} = 419.14 \text{ nm} \qquad (10.17)$$

Thus, the required difference in wavelengths is

$$\Delta\lambda = \lambda_{\text{protein1}} - \lambda_{\text{protein2}} = 423.33 - 419.14 = 4.19 \text{ nm} \qquad (10.18)$$

10.7 Magnetic Nanosensors

How has the discovery of GMR phenomenon profoundly influenced our lives? The 2007 Nobel Prize in Physics is the global recognition of the rapid development of the giant magnetoresistance (GMR), from both the physics and engineering points of view. The use of GMR can be regarded as one of the first major applications of nanotechnology (Freitas et al. 2007). GMR evoked prolific interest from researchers in fundamental physics as well as for industry. Applications of this phenomenon have revolutionized techniques for

retrieving data from hard disks. Behind the utilization of GMR structures as read heads for massive storage magnetic hard disks, breakthrough applications as solid-state magnetic nanobiosensors have emerged. The GMR nanosensor chip searches for up to 64 different proteins concomitantly and has been shown to be effective in early detection of tumors in mice, suggesting that it may open the door to much earlier detection of even the most elusive cancers in humans.

What is the significance of tunneling magnetoresistance? Magnetic tunnel junction (MTJ) sensors, which use the tunneling magnetoresistance effect (TMR), are briskly becoming the technology of choice for many magnetic sensor applications.

How do magnetic NPs act as magnetic relaxation switches (MRSws)? Magnetic NPs serve as MRSws, switching from a dispersed to a clustered state, or the reverse, due to the presence of molecular targets, with changes in the spin–spin relaxation time of water ($T2$), which can be easily detected by NMR or MRI instrumentation (Perez et al. 2004). MRSw assays are designed to cause self-assembly of magnetic NPs upon addition of molecular targets (forward switching, decreasing $T2$) or disassembly of preformed clusters by enzymatic cleavage or competitive binding (reverse switching, increasing $T2$).

Forward MRSw assays are based on cross-linking magnetic NPs into clusters using molecular target bridges. Thus, they are preferably suited for detecting small molecule analytes such as drugs, metabolites, oligonucleotides, and proteins because the short cross-links guarantee that the magnetic NPs are placed in close enough proximity to encourage relaxation switching. In addition, there is no need for time-consuming separation or capture strategies because MRSw assays can be performed in turbid solutions such as blood; removal of unbound magnetic NPs is not required. Interestingly, increasing the valency of a target by coupling to a protein or microparticle carrier has been shown to improve detection sensitivity by a factor greater than the increase in valency itself. This result is due to the evidence of higher valency targets being more effective in promoting NP clustering.

For *reverse MRSw assays*, NP clusters are first formed analogously to a forward assay before addition of an enzyme that cleaves the molecular bridges at specific sites or a competitive binding molecule that destabilize the cross-links.

How has magnetic NP labeling been used with cell surface markers? Magnetic NPs are used to tag cell surface markers so as to impart a magnetic moment to the cell that is proportional to the number of NPs bound. However, this method requires removal of excess magnetic NPs preceding the measurement of $T2$ relaxation time. But this is readily accomplished by centrifugation or filtration. The magnetic tagging concept has been commonly employed for high-contrast MR imaging.

How DNA hybridization has been detected by magnetic nanosensors? Superparamagnetic crystalline Fe_2O_3/Fe_3O_4 particles, caged in epichlorohydrin cross-linked dextran, have been used as nanosensors in DNA

hybridization experiments. The underlying principle of this approach is that the presence of a particular DNA sequence brings the particles coated with the complementary probes in close proximity, leading to an amply enhanced spin–spin relaxation of the resulting clusters.

How has peroxidase enzyme activity been sensed? Magnetic NP conjugates act as sensitive and selective nanosensors for peroxidase enzyme activity. Immobilization of dopamine or serotonin to the magnetic NPs has allowed for sensitive and direct detection by MRI of horseradish peroxidase and myeloperoxidase activity, respectively.

Example 10.5

The longitudinal/transverse relaxivity r_1 or r_2 of a maghemite contrast agent is the increase in the longitudinal/transverse relaxation rate observed in aqueous solution when the concentration in an active compound (such as iron or gadolinium) is increased by 1 mM unit. These quantities allow the comparison between different contrast agents and reflect their efficiency: the higher the relaxivities are, the smaller the quantity of agent to be injected into the patient. AMI-25 (Endorem), an SPIO (small particle of iron oxide > 40 nm) has $r_1 = 9.95\,s^{-1}\,mM^{-1}$, $r_2 = 158\,s^{-1}\,mM^{-1}$ with r_2/r_1 ratio = 15.88 while NC100150 (Clariscan), a USPIO (ultrasmall particle of iron oxide < 40 nm) has $r_1 = 24.0\,s^{-1}\,mM^{-1}$, $r_2 = 36.4\,s^{-1}\,mM^{-1}$ with r_2/r_1 ratio = 1.52. For what applications are these NPs suitable and why?

It is obvious that the transverse relaxivity of superparamagnetic contrast agents is far greater than their longitudinal relaxivity. Due to this reason, they are used mainly for $T2$-weighted imaging. SPIOs are characterized by a high r_2/r_1 ratio. However, USPIOs present a lower r_2/r_1 ratio and are therefore better suited for $T1$-weighted imaging.

10.8 Nanobiosensors

What makes QCM a useful tool in biosensing? Utilizing its ability to simultaneously detect mass and viscoelastic property changes, the QCM with dissipation monitoring (QCM-D) is the supreme tool to study biological interactions. Oligonucleotide immobilization is followed with the QCM system by measuring frequency shifts of the crystal, which was $126 \pm 12\,Hz$ under optimal condition (Duman et al. 2003). This frequency shift is sufficient for sensitive measurements.

What is the role of gold NPs in electrochemical nanosensors? Electrochemical sensing (ES) techniques are proliferating components in various fields in which an accurate, low-cost, fast, and online measuring system is indispensable. Electrochemical biosensors created by coupling biological recognition elements with electrochemical transducers based on or modified with gold

NPs have carved a niche for themselves. A few reasons explaining why gold NPs have aroused interest are as follows: (i) Au NPs provide a stable surface for the immobilization of biomolecules with retention of their biological activities, probably due to enhanced orientational freedom. This perception is extremely useful when preparing biosensors. (ii) Moreover, gold NPs allow direct electron transport between redox proteins and bulk electrode materials. Thus, they permit ES to be performed without the need for electron-transfer mediators. (iii) Desirable characteristics of gold NPs, such as their high surface-to-volume ratio, their large surface energy, their capability to decrease the distance between proteins and metal particles, and their ability to serve as electron-conducting pathways between prosthetic groups and the electrode surface, facilitates electron transfer between redox proteins and the electrode surface. (iv) The routing of electrons from redox proteins, particularly redox enzymes, to electrodes has been a subject of wide-ranging research. It is possible to electrically contact redox enzymes (GO_x) with their macroscopic environment by the reconstitution of an apo-flavoenzyme, apo-glucose oxidase, on a 1.4 nm gold nanocrystal functionalized with the cofactor flavin adenine dinucleotide.

How is DNA detection through Au NP labels enhanced? Gold NP labels, coupled with signal amplification by the silver-enhancement technique, is a proper choice of detection scheme for enhanced electrochemical detection of DNA.

Has a comparison of gold NP–decorated glucose biosensor electrodes been made? Which Au NP-tinted glucose electrode shows superior response? Comparing different amperometric glucose biosensor electrodes modified with gold NPs, the GO_x/Au_{coll}–Cyst–AuE electrode design showed a sensitivity for glucose determination higher than that achieved with GO_x/Cyst–AuE and GO_x/Au_{coll}–Cyst/Cyst–AuE and similar to that achieved with GO_x/MPA–AuE. Moreover, the useful lifetime of one single GO_x/Au_{coll}–Cyst–AuE was 28 days, remarkably longer than that of the other GO_x biosensor designs (Mena et al. 2005).

How are CNTs helpful in biosensing? For nanobiosensing applications, CNTs have multiple advantages such as small size with larger surface area, excellent electron transfer promoting ability when used as electrodes modifier in electrochemical reactions, and easy protein immobilization with retention of its activity for potential biosensors. CNTs play an important role in the performance of electrochemical biosensors, immunosensors, and DNA biosensors.

Has CNT been used as a nanoconnector? Yes, aligned reconstitution of a redox flavoenzyme (glucose oxidase) has been done on the edge of CNTs that are linked to an electrode surface. The SWCNT acts as a nanoconnector that electrically contacts the active site of the enzyme and the electrode. The electrons are transported along distances greater than 150 nm and the rate of electron transport is controlled by the length of the SWCNTs.

What is the function of Hb in a modified electrode for H_2O_2? Hb can transfer electrons directly to a normal graphite (GP) electrode when it is modified

with QDs (CdS). The Nafion/CdS–Hb/GP electrode is used as a hydrogen peroxide biosensor because of its good bioactivity.

High-impact diseases, including cancer, cardiovascular disease, and neurological disease, are exigent to diagnose without supplementing clinical evaluation with laboratory testing. *How does saliva test help in cancer detection?* A sensor array chip has been developed for direct electrochemical detection of the cancer markers (RNA and protein) from saliva associated with oral cancer (Gau and Wong 2007). The sensor assay system relies on efficient binding of target RNA molecules or proteins on to the sensor surface.

What is the contribution of aptamers in nanobiosensors? With the increasing applications of nucleic acid aptamers as a new class of molecular recognition probes in bioanalysis and biosensors, the development of general and simple signaling strategies to transduce aptamer-target binding events to detectable signals is demanding more attention. A new signaling method based on aptamers and a DNA molecular light switching complex, $[Ru(phen)_2(dppz)]^{2+}$, was developed for sensitive protein detection. Aptamer-conjugated QDs are able to specifically target U251 human glioblastoma cells with potential applications such as cancer targeting and molecular imaging. By functionalizing the surface of a QD with aptamers which recognize cocaine, and taking advantage of single-molecule detection and fluorescence resonance energy transfer between 605QD and Cy5 and Iowa Black RQ, a single-QD-based aptameric sensor was fabricated that was capable of sensing the presence of cocaine through both signal-off and signal-on modes (Zhang and Johnson 2009).

10.9 Chemical Nanosensors

Chemical sensors have been developed for decades now to detect gases and vapors at various concentration levels for deployment in a wide range of industrial applications. *In what ways was hydrogen nanosensor better than the corresponding macrosensor?* Hydrogen sensor fabricated using Si NWs modified by palladium (Pd) NPs showed superior sensitivity to hydrogen and faster response time than the macroscopic Pd wire hydrogen sensor (Chen et al. 2007).

How are CNTs used in gas detection? A sensor platform consisting of an interdigitated electrode (IDE) pattern has been fabricated for sensing gas and organic vapors. Purified SWCNTs in the form of a network laid on the IDE by solution casting serve as the sensor material. The electrical conductivity of the SWCNT network changes reproducibly upon exposure to various gases and vapors. Selectivity to specific gases, e.g., chlorine and hydrochloric acid vapor, was achieved by coating the SWCNTs with polymers such as chlorosulfonated polyethylene and hydroxypropyl cellulose (Li et al. 2006).

What is the role of metal oxide NPs in gas sensors? A semiconductor gas sensor made with a porous assembly of tiny crystals of an N-type metal oxide semiconductor, typically SnO_2, In_2O_3, or WO_3, often loaded with a small amount of foreign substance (noble metals or their oxides) called a *sensitizer,* possesses an electrical resistance. When operated at adequate temperature in air (200°C–500°C), the resistor changes its resistance sharply on contact with a small concentration of reducing gas or oxidizing gas, enabling us to know the gas concentration from the resistance change. Enhancement in selectivity and overall efficiency of the sensors is achieved by tailoring size, structure, and shape of the NPs.

What are the physicochemical models of gas sensors? Gas elements are prepared in polycrystalline form, and in the process, form large numbers of grain boundaries and necks. The model that combines the neck and the grain boundary mechanisms illustrates the effects of neck and grain boundaries on sensitivity. The gas sensitivity increases with decreasing grain size. The non-agglomerated necking of the NPs induced by the heat treatment significantly enhances the gas sensing characteristics of the NP-based gas sensors. Model of receptor function and response of small nanosize semiconductor crystals was formulated by using the chemical parameters of *gases side*, such as partial pressure, adsorption constant, and rate constant, and the physical parameters of *semiconductor side*, for example, shape and size, donor density, and Debye length.

How are CNTs used in ISFETs? A CNT-based nano ISFET was fabricated by testing of CNT *I–V* characteristic to verify if the CNT is metallic or semiconducting. The metallic CNTs are used as NWs connecting the source and drain, while the semiconducting ones are used for nanotransistors. *What are the other possible pH nanosensors?* Si NWs have been used as a sensing layer in an extended-gate field-effect transistor (EGFET) for the measurement of solution pH. Compound semiconductor/isolator (ZnS/silica) core/shell nanocables have been used to fabricate single NW-based FETs. After chemical modification, amine- and oxide-functionalized nanocables exhibit linear pH-dependent conductance.

How is moisture content of air detected by nanosensor? By coating one side of the surface of a ZnO nanobelt with multilayer polymers using an electrostatic self-assembling process, a humidity/chemical nanosensor based on piezoelectric field effect transistor (PE-FET) principle was demonstrated (Lao et al. 2007).

How do QDs help in measuring concentrations of toxic ions like antimony and mercury? A BSA-activated CdTe QD nanosensor for antimony was reported. When antimony ion enters the BSA, the QD emission is switched off (Ge et al. 2010). Luminescent and stable CdSe/ZnS core/shell QDs capped with L-carnitine were prepared for optical determination of mercury ions in ethanol via analyte-induced changes in the photoluminescence of the QDs (Li et al. 2008).

Example 10.6

Nair and Alam (2006) showed that average response settling time (t_s) and minimum detectable concentration (ρ_0) for nanobiosensors and nano-chemical sensors obey the relationship

$$\rho_0 \times t_s^{M_D} \sim k_D \tag{10.19}$$

where M_D and k_D are dimensionality dependent constants for one-, two-, and three-dimensional nanosensors. Given that $M_D = 0.5$ for planar ISFET and unity for cylindrical NW and nanosphere. Also, $k_D = N_s\sqrt{(2/D)}$ for planar ISFET where N_s is the minimum number of molecules to be captured for detection and D is the diffusion coefficient of biological or chemical target molecules in the solution. k_D is ($N_s a_0/D$) for cylindrical NW where a_0 is the radius of NW. For 12 base pair (bp) DNA, $D = 4.9 \times 10^{-6}/(bp)^{0.72}$ cm^{-2} s^{-1} at 23°C, $N_s = 10\,\mu m^{-2}$ and $a_0 = 30$ nm. Note that $N_s = 10\,\mu m^{-2}$ corresponds to 2 conjugations on a 1 μm long, 30 nm radius NW. If an experimenter waits for 5 min before recording observations, calculate the minimum detectable concentration of DNA for the planar ISFET and wire nanostructures.

Here, $D = 4.9 \times 10^{-6} / (bp)^{0.72}\,cm^2\,s^{-1}$

$$= 4.9 \times 10^{-6} / (12)^{0.72}\,cm^2\,s^{-1} = 8.1881 \times 10^{-7}\,cm^2\,s^{-1}$$

$$= 8.1881 \times 10^{-7} \times (10^{-2} \times 10^{-2})\,m^2\,s^{-1} = 8.1881 \times 10^{-11}\,m^2\,s^{-1},$$

$$N_s = 10\,\mu m^{-2} = 10 \times 10^6 \times 10^6 = 10^{13}\,m^{-2},$$

$$a_0 = 30\,nm = 30 \times 10^{-9}\,m = 3 \times 10^{-8}\,m.$$

The detection limit (ρ_0) of a planar ISFET sensor is

$$\rho_0 = \frac{k_D}{t_s^{M_D}} = \frac{N_s\sqrt{2/D}}{t_s^{0.5}} = \frac{N_s\sqrt{2/D}}{(300)^{0.5}} = \frac{10^{13} \times \sqrt{2/(8.1881 \times 10^{-11})}}{17.321}$$

$$= \frac{1.5629 \times 10^{18}}{17.321}\,\text{molecules m}^{-3} = 9.02315 \times 10^{16}\,\text{molecules m}^{-3}$$

$$= 9.02315 \times 10^{10}\,\text{molecules cm}^{-3} \tag{10.20}$$

Now, 6.022×10^{23} molecules $= 1$ mol

$$1\,\text{molecule} = 1/(6.022 \times 10^{23})\,\text{mol} = 1.661 \times 10^{-24}\,\text{mol}$$

$$9.02315 \times 10^{10} \text{ molecules} = 9.02315 \times 10^{10} \times 1.661 \times 10^{-24} \text{ mol}$$

$$= 1.498 \times 10^{-13} \text{ mol}$$

Hence, concentration $= 1.498 \times 10^{-13} \text{ mol}$ $\text{cc}^{-1} = 1.498 \times 10^{-13} \text{ mol}/10^{-3} \text{ L} = 1.498 \times 10^{-10} \text{ mol L}^{-1} = 0.1498 \text{ nM}$

The detection limit (ρ_0) of a cylindrical NW sensor is

$$\rho_0 = \frac{k_D}{t_s^{M_D}} = \frac{N_s a_0/D}{t_s^1} = \frac{\left(10^{13} \times 3 \times 10^{-8}\right)/\left(8.1881 \times 10^{-11}\right)}{300} = \frac{3.664 \times 10^{15}}{300}$$

$$= 1.2213 \times 10^{13} \text{ molecules m}^{-3} = 1.2213 \times 10^7 \text{ molecules cm}^{-3}$$

$$\text{(10.21)}$$

As before,

$$1.2213 \times 10^7 \text{ molecules} = 1.2213 \times 10^7 \times 1.661 \times 10^{-24} \text{ mol}$$

$$= 2.0286 \times 10^{-17} \text{ mol}$$

Hence, concentration $= 2.0286 \times 10^{-17} \text{ mol}$ $\text{cc}^{-1} = 2.0286 \times 10^{-17} \text{ mol}/10^{-3} \text{ L} = 2.0286 \times 10^{-14} \text{ mol L}^{-1} = 20.286 \text{ fM}$.

The ratio, detection limit of planar ISFET sensor/detection limit of NW sensor $= 0.1498 \text{ nM}/20.286 \text{ fM} = 0.1498 \times 10^{-9}/20.286 \times 10^{-15} = 7.384 \times 10^3$.

Example 10.7

In a QD fluorescence quenching analysis, the fluorescence extinction coefficient (K) was $2 \times 10^5 \text{ M}^{-1}$. If the analyte concentration was 10^{-5} M, by what percentage the fluorescence intensity might have fallen?

The extinction of fluorescence takes place according to the Stern–Volmer equation expressed as

$$\frac{I_{\max}}{I} = 1 + K \times [X] = 1 + 2 \times 10^5 \times 10^{-5} = 3 \qquad \text{(10.22)}$$

Therefore, $I/I_{\max} = 1/3 = (1/3) \times 100\% = 33.33\%$.

10.10 Nanosensor Fabrication Aspects

Looking at the enabling technologies of nanosensors, the obvious question arises: *By what approach have nanosensors been hitherto fabricated?* Although there has been momentous progress in fabrication of nanostructure-based

nanosensors, most of these functional devices and integrated systems are fabricated by a *top–down approach* using a combination of lithography, etching, and deposition. Microcontact printing, imprint lithography, or direct-write dip-pen nanolithography are different nanofabrication techniques currently being used to fabricate components with at least one of their dimensions in a scale below 100 nm. Despite several technological and physical limitations, the evolution of classical lithography techniques and other nonstandard procedures have been used in different forms to realize nanoscale components with atomic precision.

As already known, the technique for these top–down approaches is complicated, time-consuming, and expensive, which limits the practical applications to a large extent. Appreciation of the kinetics and the thermodynamics of nanostructured materials synthesized by the bottom–up and top–down approaches and their subsequent integration in sensors is an interesting research area.

What are the essential principles of bottom–up approach? In a bottom–up approach, the focus is on building smaller components into more complex assemblies. *Molecular manufacturing,* that is, the process of assembling nanodevices molecule by molecule, exemplifies a bottom–up approach. For the moment, self-assembly of nanocomponents by *DNA scaffolding* is one of the most promising techniques. It is believed that although top–down integration techniques will predominate at least for one more decade, novel bottom–up procedures, such as an evolution of DNA scaffolding, will be the way to obtain integrated nanodevices with higher complexity. *What is DNA scaffolding?* DNA molecules self-assemble in solution. This occurs via a reaction between a long single strand of viral DNA and a mixture of different short synthetic DNA strands. These short segments act as staples effectively folding the viral DNA into required two-dimensional shapes through complementary base-pair binding. Thus, DNA nanostructures such as squares, triangles, and stars are prepared measuring 100–150 nm on an edge and as thick as the DNA double helix width. Techniques have been developed to orient and position self-assembled DNA shapes and patterns, or "DNA origami," on surfaces that are well suited to semiconductor manufacturing equipment. These correctly positioned DNA nanostructures serve as scaffolds or miniature circuit boards for the precise assembly of computer-chip components.

What are the issues regarding NPs? Although some studies of the influence of shape, size, and distribution of NPs on sensing behavior, response, and recovery are reported, more work in this field is necessary. Studies on the binding of molecules to sensor elements, surface functionalization, and kinetics of adsorption and desorption of molecules will help in resolving many issues.

What are the problems related to CNTs? The "rediscovery" of CNTs in 1991 was perhaps the most noteworthy event in the short history of nanotechnology, stimulating research into all aspects of CNTs from their manufacture

to end use. The number of publications and patents on CNT synthesis is fast increasing and deluging the research journals. Still there are many remaining challenges that need to be addressed (Shanov et al. 2006). *What are these hurdles?* One of them is the production of large-scale and low-cost SWCNTs and multi-wall carbon nanotubes (MWCNT). Currently, research into nanotubes and their applications is hampered by the lack of a suitable technique for manufacturing them in large quantities, which is defined as 10,000 ton per plant per year. Concerning CNT synthesis via fluidized bed chemical vapor deposition (FBCVD), survey shows that no methodical study of the key parameters has been undertaken (See and Harris 2007). There is no clear understanding of the influence of key variables (e.g., temperature, pressure, and carbon source) on CNT properties (e.g., CNT diameter, length, and morphology), which shows that further research is necessary to optimize this process and, eventually, comprehend the science behind CNT growth using this technique. Another field of interest comprises the pursuit of controlled CNT growth in terms of selective deposition, orientation, and preselected metallic or semiconducting properties. Some of the long-standing problems in the nanotube area are due to the following reasons: (i) lack of control in the synthesis and chemical processing of SWCNTs, (ii) in chirality control that determines whether a nanotube is metallic or semiconducting, (iii) in diameter control determining the bandgap of a semiconducting SWCNT, and (iv) in the placement and orientation control on large substrates needed for scalable production of nanotube electronics and other devices (Joselevich et al. 2008).

To summarize, presently, there are four main challenges in the field of nanotube synthesis: (i) *Mass production*, that is, the development of low-cost, large-scale processes for the synthesis of high-quality nanotubes, including SWCNTs; (ii) *selective production*, that is, control over the structure and electronic properties of the manufactured nanotubes; (iii) *organization*, that is, control over the location and orientation of the produced nanotubes on a flat substrate; and (iv) *mechanism*, that is, the development of a systematic understanding of the processes of nanotube growth. Our understanding of the CNT growth mechanism has been evolving, but more consideration is still required to explain the variety of the observed growth features and experimental results.

10.11 *In Vivo* Nanosensor Problems

One of the key advantages of the small size of nanosensors is their *in vivo* use. *What are the critical issues concerning nanobiosensors to be inserted inside the human body?* For implanted sensors, the major bottleneck is *biofouling* (Drake et al. 2007). This occurs due to inflammatory response of body toward any

foreign material that is introduced in it. Various researchers have shown that this results in the decrease in sensitivity of the sensor by almost 80% as compared to sensitivity during *in vitro* use. Various types of lipids, peptides, and proteins actively get adsorbed on the biosensor and diminish its functionality.

Research is still in the process of defining the various adsorbents and possible selective adsorption of the analyte. One of the approaches that currently scientists are exploring is modifying the surface properties by *biocompatible coatings. How do biocompatible coatings function*? These coatings inhibit the binding of nonspecific elements and at the same time they do not affect the analyte. Different polymers are being used for these applications. These polymers should be biocompatible as well as non-immunogenic. Immunogenicity is the ability of a substance to provoke an immune response in the body.

What are the commonly used biocompatible polymers? Most widely used polymers for biocompatible coatings include polyethylene glycol (PEG), alkanethiols, poly(vinyl alcohol), chitosan, and poly(acrylamide) for modifying the surface.

10.12 Molecularly Imprinted Polymers for Biosensors

How can the problem of limited shelf life of biomolecules used in biosensors be solved? Biosensors utilize the advantage of the specific recognition properties of specialized biomacromolecules like antibodies or enzymes. However, the latter tend to be unstable outside their natural environment. This results in an environmental intolerance and short shelf life of the sensor. Thus, the biosensor becomes crippled after some period of time as the biomolecule loses its stability.

Synthetic biomimetic receptors such as molecularly imprinted polymers (MIPs) help to overcome these limitations. MIPs are capable of binding target molecules with affinities and selectivities akin to those of antibodies, enzymes, or hormone receptors, while being more stable, easier to prepare, less costly, and easier to integrate into standard industrial fabrication processes than the biomolecules. This is the same as aeroplanes built to mimic birds proved more efficient and powerful than the flying birds themselves, which were the genesis of the idea of flying machine.

Bompart et al. (2010) combined the specific recognition properties of MIPs with noble metal nanocomposites for signal amplification and optical readout. Based on this combination mode, they fabricated a single-particle chemical nanosensor. They synthesized molecularly imprinted core–shell nanocomposite particles by precipitation and seeded emulsion polymerization. A layer of gold colloids was located between the polymeric core and

the MIP shell. This layer resulted in signal enhancement when the particles were used as optical nanosensors. They described SERS measurements of the β-blocking drug (S)-propranolol adsorbed on to single Au-MIP core–shell composite particles of submicrometer size. By Raman measurements on single particles, the target molecule (S)-propranolol was detected after incubation at concentrations as small as 10^{-7} M.

10.13 Interfacing Issues for Nanosensors: Power Consumption and Sample Delivery Problems

Both the continued downscaling of conventional semiconductor electronics and also the next-generation "molecular-based electronics" rely profoundly on a broad array of nanoscience and technology investigations that encompass areas such as physics, chemistry, biology, material science, along with engineering sciences like electronics and mechanics, and even the computer sciences. By their very nature, all these multidisciplinary nanoscience and technology endeavors must concern themselves with molecular-level processes. They must therefore incorporate methodologies for interfacing to the microscopic phenomenon.

   The advantages of nanofabricated sensors over traditional sensors, namely, low power consumption, self-diagnosis, reliability, and cost reduction, have to be fully exploited by users. One of the most forgotten aspects in the nano-electronics field is the *problem of wiring*. Nanosensors do not have the necessary efficient interfacing tools. An effective interface to the nanosensor is needed in order to extract the embedded signals. *How does one wire individual nanoelectronic devices within a nano-integrated circuit together? Furthermore, how does one extract and input information from such a circuit,* that is, *how does one allow it to communicate with the outside world?*

   A nanoscale circuit or sensor array relinquishes its advantages of nanoscale integration if the wiring makes the final system just as heavy, bulky, or even larger than conventional systems. *What is the use if a nanosensor consumes pico or nanowatts of power while the power-hungry signal processing circuit uses milliwatts?* There will be a vast compatibility gap separating the two. For nanosensors, data acquisition and control system requirements need to be modified to be in consonance with the nanoscale sensor. Building congruous data processing and power systems is the need of the hour.

   What in our imagination is the notion of a nanosensor in actual operation? For a device such as a nanoscale sensor array, that could detect chemical and biological species as well as electrical potentials on tens of thousands of different channels, if at all possible, one would visualize a device integrated into a miniscule package that could be introduced into the human body, within the bloodstream or noninvasively in the brain. If physical wires are required

to communicate with this sensor array, its connections severely impede the potential advantages of size and mobility. In the end, one would like a system for wirelessly interfacing with the sensor array, both to extract information and provide power.

The developments of integrated circuits, which can detect, convert, process, and amplify minute signals, are required from the microelectronic/nanoelectronic community. As such, almost all the foundation work for the next generation of nanoelectronics is intrinsically defining new sensor modalities, while at the same time contributing to the advancement of the traditional capabilities (data and signal processing, computation, and communication) needed for the realization of intelligent sensors and integrated multifunctional sensor systems.

Scientists are looking forward to develop easy-to-operate systems with the ability to interface between nanoscale devices, microsystems, and macrosystems to supplement human analysis of data. The situation can be managed only by bringing novelty and through innovation of new systems. There is a demand for conceptualization and generation of new nano ideas, their implementation, and taking them to the market place.

Apart from signal interfacing, what other problem arises at the nano–micro interface? Where microtechnology meets nanotechnology, not only electrical interfacing problems arise, the delivery of microliter samples of analytes done by microfluidic devices is no longer acceptable because nanoliters of samples are required for which suitable nanofluidic devices must be fabricated. Thus, the transition from micro to nanoworld must be envisaged in its entirety and this aspect of interfacing, notably, sample delivery, should not be ignored. The sample quantities used in microfluidics will inundate the nanosensor, which is neither necessary nor conducive to sensor functioning. Exactly similar situation arose when microsensors were developed. The obvious question is as follows: *if the sensor does not need microliters of blood sample, why should the same be withdrawn from a patient's body?*

10.14 Depletion-Mediated Piezoelectric Actuation for NEMS

What is the concept of depletion-mediated piezoelectric actuation? Masmanidis et al. (2007) investigated the use of piezoelectric semiconductor in their nanomechanical device. The device consists of a 200 nm thick epitaxially grown GaAs P-type/intrinsic/N-type (PIN) diode. An N-doped layer serves as the top electrode, and finally, the *charge depleted high-resistance region* in the middle forms the piezoelectrically active layer. They used NEMS cantilevers to demonstrate and estimate the efficiency of depletion-mediated NEMS piezoelectric excitation. During a classic measurement, an AC signal applied across the PIN junction actuated the device at or near its resonance frequency, while

the addition of DC voltage tuned the depletion region width. The nanoscale dimensions of the structure concentrated the electric field across a small width, which imparted excellent voltage sensitivity. Within the device, a transverse electric field, E_z, produced a longitudinal strain.

How does the traditional view change at the nanoscale? The traditional view calls for well-defined, alternating layers of electrodes and piezoelectric materials. But this view crumbles in nanoscale devices made from semiconductors, where charge depletion smears out the boundary between piezoelectrically inactive (electrically conducting) and active (electrically insulating) regions. This spreading effect is only significant when the total device thickness approaches the charge depletion width of the semiconductor. A bending moment was developed when the strain was asymmetrically distributed around the neutral axis of the beam. This bending moment resulted in mechanical resonance under a suitable range of driving frequencies.

The piezoelectric effect was capable of driving the device with AC signals as low as 5 mV before the inception of thermomechanical fluctuations, corresponding to roughly speaking, a single electronic charge on the cantilever itself. *What was the power consumption?* Assuming a maximum current flow of 1 nA at 5 mV of AC drive, the minimum required power consumption of this device approached 5 fW, with ~1 nW being more representative of typical operating conditions (an AC drive of 10 mV) during actuation. The reversibility of piezoelectric phenomena offered the potential for ultrasensitive electrical measurement of nanomechanical motion.

Thus, they employed epitaxial piezoelectric semiconductors to obtain efficient and fully integrated NEMS actuation, which was based on exploitation of the interaction between piezoelectric strain and built-in charge depletion. Other NEMS devices needed extrinsic methods of actuation, making the entire system bulky. In addition, they showed that the devices could be tuned in frequency using a DC bias, making them sensitive detectors of electrical potentials. The nanoscale dimensions of the structure concentrated the electric field across a small width, which imparted excellent voltage sensitivity. This tuning effect was linear, unlike other NEMS devices which are nonlinear in response. Finally, since all of the concepts presented by above workers were extendable to a wide variety of other materials beyond GaAs (such as AlN, SiC, or ZnO), enhanced electrical and mechanical properties are a natural and irresistible consequence of this actuation method.

10.15 Discussion and Conclusions

Nanoscience and nanotechnology deal with the study and application of structures of matter with at least one dimension of the order of less than 100 nm. However, *properties related to low dimensions are more important than size.*

Nanotechnology is based on the fact that some very small structures usually have new properties and behavior that are not displayed by the bulk matter with the same composition. With the advent of nanotechnology, research is underway to create miniaturized sensors. Miniaturized sensors can lead to reduced weight, batch fabrication, lower power consumption, and low cost.

Research activity in the areas related to nanosensors has seen phenomenal growth in the last one-and-half decade. For well over a decade, nanoscale science and technology (nano-S&T) related subjects have represented an exponentially growing portion of nearly all scientific and engineering disciplines. Indeed, the possibility of leveraging fundamental mechanisms at the nanoscale and molecular levels through a progressively increasing capability to understand, prescribe, and control all the basic properties (structural, chemical, mechanical, electronic, and photonic) of ultrasmall systems seems to offer fantastic and unlimited new possibilities for the future. In this book, an attempt was made to provide the most contemporary overview possible of nanosensors and their potential applications. Several nanostructures that are currently used in the development of nanosensors: NPs, nanotubes, nanorods, porous silicon, and self-assembled materials, were described, focusing on the type of nanomaterial used and the properties related to the particular nanostructure.

But all that has been said is just a drop in the ocean! Nanosensors promise reaching far ahead of microsensors, for example, going a further step forward of the microarray technology, nanoarrays are being developed based on the interaction between different types of receptors and ligands such as proteins or nucleic acids; approximately 400 nanoarray spots can be placed in the same area as a traditional microarray spot (Nanosensors in environmental analysis 2006: http://www.nanowerk.com/spotlight/spotid=366.php).

"Small is beautiful." Today we know that small is not only beautiful, but also powerful, looking at the range of applications that nanosensors find in varied fields of engineering and technology. Nanosensors represent a new frontier in technology, blurring the lines and defining a sort-of conjugated system between physics, chemistry, biology, and materials. Essential and desirable features of nanosensors are summarized in the following lines:

> *All dimensions of nanosensors need not to be nanoscale,*
> *But they work only where "nanostate properties" prevail:*
> *Quantum-mechanical tunneling, Quantum confinement effect,*
> *Giant magnetoresistance, LSPR, SERS, QD-FRET,…*
> *That is the nanosensor definition,*
> *For everyone's attention!*
> *What if all their dimensions be in nanorange?*
> *Tinnier and better but still nanosensor; status unchanged!*
> *Sensitivities of nanosensors lie in the "nanoregime,"*
> *The greater the sensitivity, the higher their esteem.*
> *Their power consumption: nanowatts of power,*
> *Lesser power-hungry devices are need of the hour.*

They are batch-fabricated in a large foundry,
Still are low-priced, and almost free…
But interfacing them to the macroworld,
Is everyone's puzzle.
Building efficient nanogenerators is an arduous job,
…Behold, our hearts throb!
For the future, nanosensors have potentials immense,
Much remains unexplored, with lots of suspense,
Frenzy and excitement!
In nanosensor research and development.
Nanosensors: Physical, Chemical and Biological,
Will make us happy, healthy and comfortable.
Bridging ideas into practical realization,
Nanosensors will shrink instrumentation,
And write a new page!
Of glorious nanotechnology age!

The material presented in the chapters of this book was intended to serve only as a foundation course. Detailed information can be found in the excellent references provided at the end of each chapter, which the eager reader is recommended to kindly peruse in earnest.

Review Exercises

10.1   The cantilever structure has been widely used for fabricating nanosensors. Prepare a status review of the use of cantilevers in various types of nanosensors.

10.2   CNTs are sensitive to different physical and chemical properties. Compile a table illustrating the huge variety of nanosensors that utilize CNTs.

10.3   Metallic NPs have been used in nanosensors for performing various tasks. List and elucidate the different roles of NPs in building nanosensors.

10.4   QDs are used in optical nanosensors. They are also used in electrochemical nanosensors. Elucidate these two widely different applications of QDs as nanosensors.

10.5   Describe a model for the functioning of a gas sensor and discuss the influence of NP shape on sensitivity.

10.6   Explain the relevance of diffusion of gas into the porous sensing material toward its sensing characteristics. Under steady-state conditions, the gas concentration inside the sensing layer decreases with depth

resulting in a diffusion-controlled gas concentration profile. What is the effect of such variation in penetration of gas in the sensing layer, on the sensitivity of the sensor, and its response transients?

10.7 Arrhenius equation describes the temperature dependence of the conductivity of a semiconductor: The conductivity is given by

$$\sigma = \sigma_0 \exp\left(-\frac{qV_s}{kT}\right) \tag{10.23}$$

where

σ_0 is a pre-exponential factor that includes the bulk intragranular conductance

k is the Boltzmann's constant

T is the absolute temperature

qV_s is the potential energy barrier at the interface between two neighboring particles

The potential energy barrier is expressed as

$$qV_s = \frac{q^2 N_t^2}{2\varepsilon_0 \varepsilon_s N_d} \tag{10.24}$$

where

N_t is the surface density of adsorbed oxygen ions

ε_0, ε_s are the permittivity of free space and the semiconductor, respectively

N_d is the volumetric density of the electron donors

Apply Equations 10.23 and 10.24 to discuss the effect of particle size on the properties of a gas sensor, especially when particle size is reduced to nanometers (see Zhang and Liu 2000).

10.8 An experiment is done with nanostructured silica shells of diatoms: a type of algae, as follows: laser beams are shone at the shells of *Thalassiosira rotula* in the presence of nitrous oxide, acetone, ethanol, air, xylene, and pyridine. The wavelengths of the light emitted by the shells are measured.

Are these wavelengths the same? Do the shells present slightly different colors depending on the surrounding gas? What is the application implied?

10.9 pH sensitivity of silicon bulk materials is poor. However, a pH sensitivity of 58.3 mV pH^{-1} was observed for Si NW. Why?

10.10 Is GMR unique to multilayered structures? Can GMR occur in magnetically inhomogeneous media? (See Xiao et al. 1992.)

10.11 How are magnetic particles prevented from self-aggregation? What parameter changes when the particles assemble together to form nanoclusters? How is this property utilized in making a nanosensor?

10.12 Describe a magnetic nanosensor technology that is matrix-insensitive, yet capable of rapid, multiplex protein detection with resolution down to attomolar concentrations and extensive linear dynamic range.

10.13 Explain the term "biofouling" in reference to a biosensor. How can it be prevented?

10.14 What is the application of MIPs in biosensors? Give one example of their use.

10.15 The use of conventional approaches such as magnetomotive, electrostatic, and electrothermal techniques for NEMS actuation, suffer from either low power efficiency, limited potential for integration, or poor nanoscale control over electromechanical coupling. Explain. How does piezoelectric effect help in NEMS actuation?

10.16 How does the piezoelectric effect inherent to a material afford a highly efficient means of resonantly exciting NEMS devices with an alternating voltage? What is the concept of tunably coupled NEMS actuation?

10.17 Argue that nanosensors must be wireless or small battery-operated devices to utilize the real benefit of their small size.

References

Akyildiz, I. F. and J. M. Jornet. 2010. Electromagnetic wireless nanosensor networks. *Nano Communication Networks* 1: 3–19.

Barbillon, G., J.-L. Bijeon, J. Plain, M. L. de la Chapelle, P.-M. Adam, and P. Royer. 2007. Biological and chemical gold nanosensors based on localized surface plasmon resonance. *Gold Bulletin* 40(3): 240–244.

Bompart, M., Y. D. Wilde, and K. Haupt. 2010. Chemical nanosensors based on composite molecularly imprinted polymer particles and surface-enhanced Raman scattering. *Advanced Materials* 22: 2343–2348.

Bourlon, B., J. Wong, C. Miko, L. Forró, and M. Bockrath. 2007. A nanoscale probe for fluidic and ionic transport. *Nature Nanotechnology* 2: 104–107.

Chai, G., O. Lupan, L. Chow, and H. Heinrich. 2009. Crossed zinc oxide nanorods for ultraviolet radiation detection. *Sensors and Actuators A* 150: 184–187.

Chen, Z. H., J. S. Jie, L. B. Luo, H. Wang, C. S. Lee, and S. T. Lee. 2007. Applications of silicon nanowires functionalised with palladium nanoparticles in hydrogen sensors. *Nanotechnology* 18: 345502 (5pp.), doi: 10.1088/0957-4484/18/34/345502

Chovin, A., P. Garrigue, G. Pecastaings, H. Saadaoui, and N. Sojic. 2006. Development of an ordered microarray of electrochemiluminescent nanosensors. *Measurement Science and Technology* 17: 1211–1219.

Chow, W. W. Y., Y. Qu, W. J. Li, and S. C. H. Tung. 2010. Integrated SWCNT sensors in micro-wind tunnel for air-flow shear-stress measurement. *Microfluidics and Nanofluidics* 8: 631–640, doi: 10.1007/s10404-009-0495-5

Drake, C., S. Deshpande, D. Bera, and S. Seal. 2007. Metallic nanostructured materials based sensors. *International Materials Reviews* 52(5): 289–317.

Duman, M., R. Saber and E. Pi kin. 2003. A new approach for immobilization of oligonucleotides onto piezoelectric quartz crystal for preparation of a nucleic acid sensor for following hybridisation. *Biosensors and Bioelectronics* 18(11): 1355–1363.

Ekinci, K. L. 2005. Electromechanical transducers at the nanoscale: Actuation and sensing of motion in nanoelectromechanical systems (NEMS). *Small* 1(8–9): 786–797.

Freitas, P. P., R. Ferreira, S. Cardoso, and F. Cardoso. 2007. Magnetoresistive sensors. *Journal of Physics: Condensed Matter* 19: 165221, doi: 10.1088/0953-8984/19/16/165221

Gau, V. and D. Wong. 2007. Oral fluid nanosensor test (OFNASET) with advanced electrochemical-based molecular analysis platform. *Annals of the New York Academy of Sciences* 1098: 401–410, doi: 10.1196/annals.1384.005

Ge, S., C. Zhang, Y. Zhu, J. Yu, and S. Zhang. 2010. BSA activated CdTe quantum dot nanosensor for antimony ion detection. *Analyst* 135: 111–115.

Giessibl, F. J. 2003. Advances in atomic force microscopy. *Reviews of Modern Physics* 75: 949–983.

Joselevich, E., H. Dai, J. Liu, K. Hata, and A. H. Windle. 2008. Carbon nanotube synthesis and organization. *Topics in Applied Physics* 111: 101–164.

Khanna, V. K. 2008. Nanoparticle-based sensors. *Defence Science Journal:* Special Issue on Nanomaterials: Science and Technology-II, 58(5): 608–616.

Khanna, V. K. 2009. Frontiers of nanosensor technology. *Sensors and Transducers Journal* 103(4): 1–16.

Kneipp, J., B. Wittig, H. Bohr, and K. Kneipp. 2010. Surface-enhanced Raman scattering: A new optical probe in molecular biophysics and biomedicine. *Theoretical Chemistry Accounts* 125: 319–327.

Knobel, R. G. and A. N. Cleland. 2003. Nanometre-scale displacement sensing using a single electron transistor. *Nature* 424: 291–293.

Lao, C. S., Q. Kuang, Z. L. Wang, M.-C. Park, and Y. Deng. 2007. Polymer functionalized piezoelectric-FET as humidity/chemical nanosensors. *Applied Physics Letters* 90: 262107-1–262107-3.

Li, J., Y. Lu, and M. Meyyappan. 2006. Nano chemical sensors with polymer-coated carbon nanotubes. *IEEE Sensors Journal* 6(5): 1047–1051.

Li, H., Y. Zhang, X. Wang, and Z. Gao. 2008. A luminescent nanosensor for Hg(II) based on functionalized CdSe/ZnS quantum dots. *Microchimica Acta* 160: 119–123, doi: 10.1007/s00604-007-0816-x

Masmanidis, S. C., R. B. Karabalin, I. D. Vlaminck, G. Borghs, M. R. Freeman, and M. L. Roukes. 2007. Multifunctional nanomechanical systems via tunably coupled piezoelectric actuation. *Science* 317: 780–782.

Mena, M. L., P. Yáñez-Sedeño, and J. M. Pingarrón. 2005. A comparison of different strategies for the construction of amperometric enzyme biosensors using gold nanoparticle-modified electrodes. *Analytical Biochemistry* 336: 20–27.

Nair, P. R. and M. A. Alam. 2006. Performance limits of nanobiosensors. *Applied Physics Letters* 88: 233120-1–233120-3.

Ozasa, A., R. Kometani, T. Morita, K. Kondo, K. Kanda, Y. Haruyama, J. Fujita, T. Kaito, and S. Matsui. 2004. Fabrication and evaluation of thermal nanosensor by focused-ion-beam chemical-vapor-deposition. In: *Microprocesses and Nanotechnology Conference*, Osaka, Japan, October 27–29, 2004, 28P-7-15, pp. 266–267.

Perez, J. M., L. Josephson, and R. Weissleder. 2004. Use of magnetic nanoparticles as nanosensors to probe for molecular interactions. *ChemBioChem* 5: 261–264, doi: 10.1002/cbic.200300730

See, C. H. and A. T. Harris. 2007. A review of carbon nanotube synthesis via fluidized-bed chemical vapor deposition. *Industrial & Engineering Chemistry Research* 46(4): 997–1012, doi: 10.1021/ie060955b

Shanov, V., Y.-H. Yun, and M. J. Schulz. 2006. Synthesis and characterization of carbon nanotube materials (review). *Journal of the University of Chemical Technology and Metallurgy* 4(4): 377–439.

Sumner, J. P., N. M. Westerberg, A. K. Stoddard, C. A. Fierke, and R. Kopelman. 2006. Cu^+- and Cu^{2+}-sensitive PEBBLE fluorescent nanosensors using DsRed as the recognition element. *Sensors and Actuators B* 113: 760–767.

Wei, J., D. Olaya, B. S. Karasik, S. V. Pereverzev, A. V. Sergeev, and M. E. Gershenson. 2008. Ultrasensitive hot-electron nanobolometers for terahertz astrophysics. *Nature Nanotechnology* 3: 496–500.

Xiao, J. Q., J. S. Jiang, and C. L. Chien. 1992. Giant magnetoresistance in nonmultilayer magnetic systems. *Physical Review Letters* 68(25): 3749–3752.

Zhang, C.-Y. and L.W. Johnson. 2009. Single quantum-dot-based aptameric nanosensor for cocaine. *Analytical Chemistry* 81: 3051–3055.

Zhang, G. and M. Liu. 2000. Effect of particle size and dopant on properties of SnO_2-based gas sensors. *Sensors and Actuators B* 69: 144–152.