1

Generic methodologies for nanotechnology: classification and fabrication

1.1 INTRODUCTION AND CLASSIFICATION

1.1.1 What is nanotechnology?

Nanotechnology is the term used to cover the design, construction and utilization of functional structures with at least one characteristic dimension measured in nanometres. Such materials and systems can be designed to exhibit novel and significantly improved physical, chemical and biological properties, phenomena and processes as a result of the limited size of their constituent particles or molecules. The reason for such interesting and very useful behaviour is that when characteristic structural features are intermediate in extent between isolated atoms and bulk macroscopic materials; i.e., in the range of about 10^{-9} m to 10^{-7} m (1 to 100 nm), the objects may display physical attributes substantially different from those displayed by either atoms or bulk materials. Ultimately this can lead to new technological opportunities as well as new challenges.

1.1.2 Classification of nanostructures

As we have indicated above, a reduction in the spatial dimension, or confinement of particles or quasiparticles in a particular crystallographic direction within a structure generally leads to changes in physical properties of the system in that direction. Hence one classification of nanostructured materials and systems essentially depends on the number of dimensions which lie within the nanometre range, as shown in Figure 1.1: (a) systems confined in three dimensions, (b) systems confined in two dimensions, (c) systems confined in one dimension.

Nanoscale Science and Technology Edited by R. W. Kelsall, I. W. Hamley and M. Geoghegan

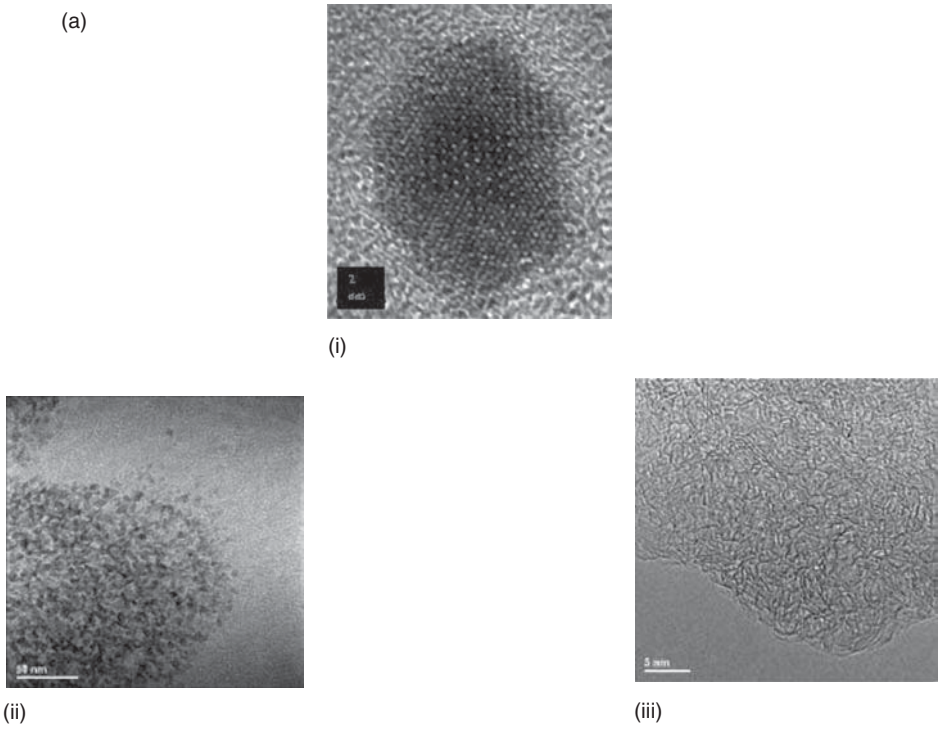


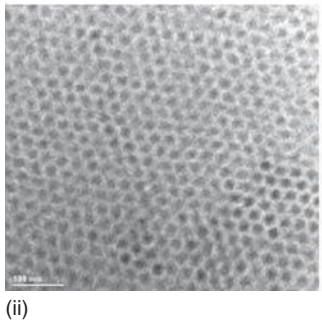
Figure 1.1 Classification of nanostructures. (a) Nanoparticles and nanopores (nanosized in three dimensions): (i) high-resolution TEM image of magnetic iron oxide nanoparticle, (ii) TEM image of ferritin nanoparticles in a liver biopsy specimen, and (iii) high-resolution TEM image of nanoporosity in an activated carbon). (b) Nanotubes and nanofilaments (nanosized in two dimensions): (i) TEM image of single-walled carbon nanotubes prepared by chemical vapour deposition, (ii) TEM image of ordered block copolymer film, and (iii) SEM image of silica nanotube formed via templating on a tartaric acid crystal. (c) Nanolayers and nanofilms (nanosized in one dimension): (i) TEM image of a ferroelectric thin film on an electrode, (ii) TEM image of cementite (carbide) layers in a carbon steel, and (iii) high-resolution TEM image of glassy grain boundary film in an alumina polycrystal. Images courtesy of Andy Brown, Zabeada Aslam, Sarah Pan, Manoch Naksata and John Harrington, IMR, Leeds

Nanoparticles and nanopores exhibit three-dimensional confinement (note that historically pores below about 100 nm in dimension are often sometimes confusingly referred to as micropores). In semiconductor terminology such systems are often called quasi-zero dimensional, as the structure does not permit free particle motion in any dimension.

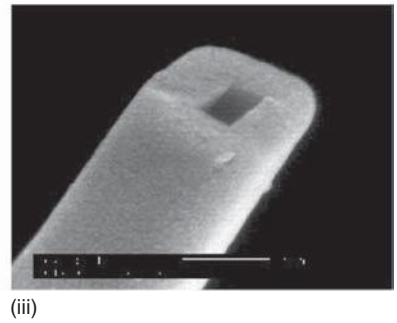
Nanoparticles may have a random arrangement of the constituent atoms or molecules (e.g., an amorphous or glassy material) or the individual atomic or molecular units may be ordered into a regular, periodic crystalline structure which may not necessarily be the same as that which is observed in a much larger system (Section 1.3.1). If crystalline, each nanoparticle may be either a single crystal or itself composed of a number of different crystalline regions or grains of differing crystallographic orientations (i.e., polycrystalline) giving rise to the presence of associated grain boundaries within the nanoparticle.

(b)

(i)

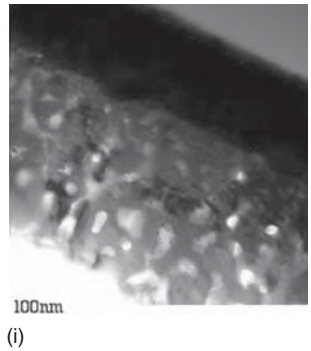


(ii)

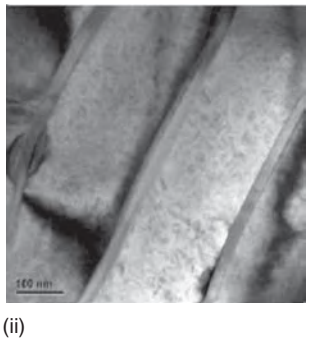


(iii)

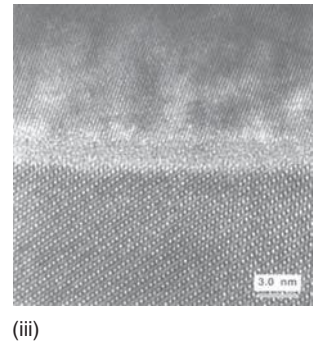
(c)



(i)



(ii)



(iii)

Figure 1.1 Continued

Nanoparticles may also be quasi-crystalline, the atoms being packed together in an icosahedral arrangement and showing non-crystalline symmetry characteristics. Such quasi-crystals are generally only stable at the nanometre or, at most, the micrometre scale.

Nanoparticles may be present within another medium, such as nanometre-sized precipitates in a surrounding matrix material. These nanoprecipitates will have a specific morphology (e.g., spherical, needle-shaped or plate-shaped) and may possess certain crystallographic orientation relationships with the atomic arrangement of the matrix depending on the nature (coherency) of the interface which may lead to coherency strains in the particle and the matrix. One such example is the case of self-assembled semiconductor quantum dots, which form due to lattice mismatch strain relative to the surrounding layers and whose geometry is determined by the details of the strain field (Chapter 3). Another feature which may be of importance for the overall transport properties of the composite system is the connectivity of such nanometre-sized regions or, in the case of a nanoporous material, nanopore connectivity.

In three dimensions we also have to consider collections of consolidated nanoparticles; e.g., a nanocrystalline solid consisting of nanometre-sized crystalline grains each in a specific crystallographic orientation. As the grain size d of the solid decreases the proportion of atoms located at or near grain boundaries, relative to those within the interior of a crystalline grain, scales as $1/d$. This has important implications for properties in ultrafine-grained materials which will be principally controlled by interfacial properties rather than those of the bulk.

Systems confined in two dimensions, or quasi-1D systems, include nanowires, nanorods, nanofilaments and nanotubes: again these could either be amorphous, single-crystalline or polycrystalline (with nanometre-sized grains). The term 'nanoropes' is often employed to describe bundles of nanowires or nanotubes.

Systems confined in one dimension, or quasi-2D systems, include discs or platelets, ultrathin films on a surface and multilayered materials; the films themselves could be amorphous, single-crystalline or nanocrystalline.

Table 1.1 gives examples of nanostructured systems which fall into each of the three categories described above. It can be argued that self-assembled monolayers and multi layered Langmuir–Blodgett films (Section 1.4.3.1) represent a special case in that they represent a quasi-2D system with a further nanodimensional scale within the surface film caused by the molecular self-organization.

1.1.3 Nanoscale architecture

Nanotechnology is the design, fabrication and use of nanostructured systems, and the growing, shaping or assembling of such systems either mechanically, chemically or biologically to form nanoscale architectures, systems and devices. The original vision of Richard Feynman[1] was of the 'bottom-up' approach of fabricating materials and devices at the atomic or molecular scale, possibly using methods of self-organization and self-assembly of the individual building blocks. An alternative 'top-down' approach is the

[1] R. Feynman, There's plenty of room at the bottom, *Eng. Sci.* **23**, 22 (1960) reprinted in *J. Micromech Systems* **1**, 60 (1992).

Table 1.1 Examples of reduced-dimensionality systems

3D confinement
Fullerenes
Colloidal particles
Nanoporous silicon
Activated carbons
Nitride and carbide precipitates in high-strength low-alloy steels
Semiconductor particles in a glass matrix for non-linear optical components
Semiconductor quantum dots (self-assembled and colloidal)
Quasi-crystals
2D confinement
Carbon nanotubes and nanofilaments
Metal and magnetic nanowires
Oxide and carbide nanorods
Semiconductor quantum wires
1D confinement
Nanolaminated or compositionally modulated materials
Grain boundary films
Clay platelets
Semiconductor quantum wells and superlattices
Magnetic multilayers and spin valve structures
Langmuir–Blodgett films
Silicon inversion layers in field effect transistors
Surface-engineered materials for increased wear resistance or corrosion resistance

ultraminiaturization or etching/milling of smaller structures from larger ones. These methods are reviewed in Section 1.4. Both approaches require a means of visualizing, measuring and manipulating the properties of nanostructures; computer-based simulations of the behaviour of materials at these length scales are also necessary. This chapter provides a general introduction to the preparation and properties of nanostructures, whilst the subsequent chapters give greater detail on specific topics.

1.2 SUMMARY OF THE ELECTRONIC PROPERTIES OF ATOMS AND SOLIDS

To understand the effects of dimensionality in nanosystems, it is useful to review certain topics associated with the constitution of matter, ranging from the structure of the isolated atom through to that of an extended solid.

1.2.1 The isolated atom

The structure of the atom arises as a direct result of the wave–particle duality of electrons, which is summarized in the de Broglie relationship, $\lambda = h/m_e v$, where λ is the (electron) wavelength, m_e is the (electron) mass, v is the velocity and

$h = 6.63 \times 10^{-34}$ J s is the Planck constant. The wave–particle duality of the electron means that an electron behaves both as a wave (i.e., it is extended over space and has a wavelength and hence undergoes wave-like phenomena such as diffraction) and a particle (i.e., it is localized in space and has a position, a velocity and a kinetic energy). This is conveniently summarized in the idea of a wave packet a localized wave that is effectively the summation of a number of different waves of slightly differing wavelengths.

Using these ideas we come to our first model of the atom, the Rutherford–Bohr model. Here the small central nucleus of the atom consists of positively charged protons and (neutral) neutrons. Electrons orbit the nucleus in stable orbits. The allowed, stable orbits are those in which the electron wavelength, given by the de Broglie formula, is an integral multiple n of the circumference of the orbit r:

$$2\pi r = n\lambda = \frac{nh}{m_e v}. \tag{1.1}$$

This implies that

$$m_e v r = \frac{nh}{2\pi}, \tag{1.2}$$

in otherwords, the angular momentum $m_e vr$ is quantized in that it is an integral multiple of $h/2\pi$.

The Bohr model leads to the idea that only certain electron orbits or shells are allowed by this quantization of angular momentum (i.e., the value of n). The Bohr shells in an atom are labelled according to the quantum number, n, and are given the spectroscopic labels K, L, M, N, etc. (where $n = 1, 2, 3, 4, \ldots$). To understand the form of the periodic table of elements, it is necessary to assume that each Bohr shell can contain $2n^2$ electrons. For instance, a K shell ($n = 1$) can contain 2 electrons, whereas an L shell ($n = 2$) can accommodate 8 electrons. As well as having a distinct form and occupancy, each shell also has a corresponding well-defined energy. It is usual to define the zero of the energy scale (known as the vacuum level) as the potential energy of a free electron far from the atom. In order to correspond with atomic emission spectra measured experimentally, the energies of these levels E_n are then negative (i.e., the electrons are bound to the atom) and are proportional to $1/n^2$. Such a simplified picture of the structure of an isolated Mg atom and the associated energy level diagram are shown in Figure 1.2.

A much more sophisticated model of the atom considers the wave-like nature of the electrons from the very beginning. This uses wave mechanics or quantum mechanics.

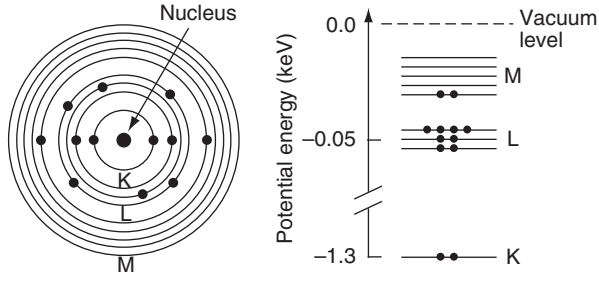


Figure 1.2 Bohr shell description of an Mg atom and the associated energy level diagram

Here each electron is described by a wavefunction ψ which is a function of spatial position (x,y,z) and, in general, of time. Physically $|\psi|^2$ represents the probability of finding the electron at any point. To work out the energy of each electron, we need to solve the Schrödinger equation which, in the time-independent case, takes the form

$$-\frac{\hbar^2}{2m_e}\nabla^2\psi + V(x,y,z)\psi = E\psi\,, \tag{1.3}$$

where V(x,y,z) describes the potential energy function in the environment of the electron. Solution of the Schrödinger equation, under certain boundary conditions, leads to a set of solutions for the allowed wavefunctions ψ_n of the atomic electrons together with their associated energies E_n.

This equation can only be solved analytically for the case of the hydrogen atom, where there is only one electron moving in the potential of a single proton, the hydrogen nucleus. Only a certain set of electronic wavefunctions and associated energy levels fulfil this Schrödinger equation. The wavefunctions may be expressed as a radial part, governing the spatial extent of the wavefunction, multiplied by a spherical harmonic function which determines the shape. The allowed wavefunctions form the electron orbitals, which we term 1s, 2s, 2p, 3s, 3p, 3d, etc. (here $1, 2, 3, \ldots$ are alternative labels for $\mathrm{K}, \mathrm{L}, \mathrm{M}, \ldots$). These allowed wavefunctions now depend on not just one quantum number but four: n, l, m and s. These numbers may be summarized as follows:

- n is the principal quantum number; it is like the quantum number used for the case of Bohr shells ($n = 1, 2, 3, \ldots$).
- l is the angular momentum quantum number; it can vary from $l = 0, 1, 2, \ldots, (n-1)$. The value of l governs the orbital shape of the subshell: $l = 0$ is an s orbital, which is spherical; $l = 1$ is a p orbital, which has a dumbbell shape; while $l = 2$ is a d orbital, which has a more complex shape such as a double dumbbell.
- m is the magnetic quantum number; it can vary from $m = 0, \pm1, \ldots, \pm l$. The value of m governs the spatial orientation of the different orbitals within a subshell; i.e., there are three p orbitals ($l = 1$) p_x, p_y, and p_z corresponding to the three values of m which are 0, +1 and −1. In the absence of a magnetic field, all these orbitals within a particular subshell will have the same energy.
- s is the spin quantum number which, for an electron, can take the values $\pm1/2$. Each (n, l, m) orbital can contain two electrons of opposite spin due to the Pauli exclusion principle, which states that no two electrons can have the same four quantum numbers.

Using this identification in terms of the quantum numbers, each electron orbital in an atom therefore has a distinct combination of energy, shape and direction (x, y, z) and can contain a maximum of two electrons of opposite spin.

In an isolated atom, these localized electronic states are known as Rydberg states and may be described in terms of simple Bohr shells or as combinations of the three quantum numbers n, l and m known as electron orbitals. The Bohr shells (designated $\mathrm{K}, \mathrm{L}, \mathrm{M}, \ldots$) correspond to the principal quantum numbers n equal to 1, 2, 3, etc. Within each of these shells, the electrons may exist in $(n-1)$ subshells (i.e., s, p, d, or f subshells, for which the angular momentum quantum number l equals 0, 1, 2, 3, respectively).

The occupation of the electronic energy levels depends on the total number of electrons in the atom. In the hydrogen atom, which contains only one electron, the set of Rydberg states is almost entirely empty except for the lowest-energy 1s level which is half full. As we go to higher energies, the energy spacing between these states becomes smaller and smaller and eventually converges to a value known as the vacuum level ($n = \infty$), which corresponds to the ionization of the inner-shell electron. Above this energy the electron is free of the atom and this is represented by a continuum of empty electronic states. In fact, the critical energy to ionize a single isolated hydrogen atom is equal to 13.61 eV and this quantity is the Rydberg constant.

This description is strictly only true for hydrogen; however, other heavier atoms are found to have similar wavefunction (hydrogenic-like) solutions, which ultimately leads to the concept of the periodic table of elements, as each atom has more and more electrons which progressively fill the allowed energy levels. This is shown for a magnesium atom in Figure 1.2. The chemical properties of each atom are then principally determined by the number of valence electrons in the outermost electron shell which are relatively loosely bound and available for chemical reaction with other atomic species.

1.2.2 Bonding between atoms

One way to picture the bonding between atoms is to use the concept of Molecular Orbital (MO) Theory. MO theory considers the electron wavefunctions of the individual atoms combining to form molecular wavefunctions (or molecular orbitals as they are known). These orbitals, which are now delocalized over the whole molecule, are then occupied by all the available electrons from all the constituent atoms in the molecule. Molecular orbitals are really only formed by the wavefunctions of the electrons in the outermost shells (the valence electrons); i.e., those which significantly overlap in space as atoms become progressively closer together; the inner electrons remain in what are essentially atomic orbitals bound to the individual atoms.

A simple one-electron molecule is the H_2^+ ion, where we have to consider the interactions (both attractive and repulsive) between the single electron and two nucleii. The Born–Oppenheimer approximation regards the nuclei as fixed and this simplifies the Hamiltonian used in the Schrödinger equation for the molecular system. For a one-electron molecule, the equation can be solved mathematically, leading to a set of molecular wavefunctions ψ which describe molecular orbitals and depend on a quantum number λ which specifies the angular momentum about the internuclear axis. Analogous to the classification of atomic orbitals (AOs) in terms of angular momentum l as s, p, d, etc., the MOs may be classified as σ, π, δ depending on the value of λ ($\lambda = 0, 1, 2$, respectively). Very simply a σ MO is formed from the overlap (actually a linear combination) of AOs parallel to the bond axis, whereas a π MO results from the overlap of AOs perpendicular to the bond axis. For the H_2^+ ion, the two lowest-energy solutions are known as $1s\sigma_g$ and $1s\sigma_u$. Here 1s refers to the original atomic orbitals; the subscripts g and u refer to whether the MO is either symmetrical or non-symmetrical with respect to inversion about a line drawn between the nucleii (viz. an even or odd mathematical function). This is shown in figure 1.3.

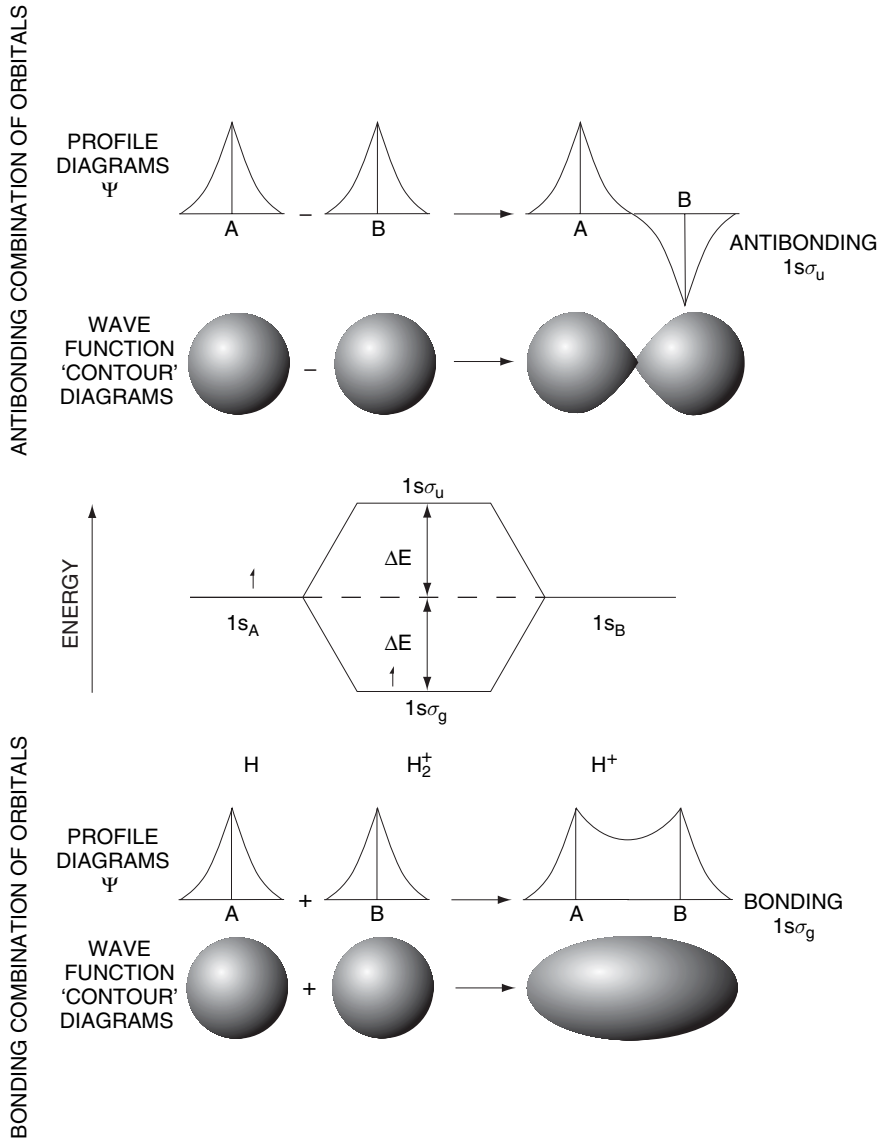


Figure 1.3 Molecular orbital description and energy level diagram for an H_2^+ ion

As can be seen the electron density is concentrated between the nuclei for the $1s\sigma_g$ MO, which is known as a bonding orbital since the energy of the molecular wavefunction is lower (i.e., more stable) than the corresponding isolated atomic wavefunctions. Conversely, the electron density is diminished between the nuclei for $1s\sigma_u$, which is known as an antibonding orbital since the energy of the molecular wavefunction is higher (i.e., less stable) than the corresponding isolated atomic wavefunction.

More generally, it is necessary to be able to solve the Schrödinger equation for molecules containing more than one electron. One way to do this is to use approximate

solutions similar to those obtained for the hydrogen atom, since when an electron is near a particular nucleus it will have a hydrogen-like form. Using this approach we can then construct a set of molecular orbitals from a linear combination of atomic orbitals (LCAO). For instance, as shown in Figure 1.4, the $1s\sigma_g$ bonding MO is formed from the in-phase overlap (i.e., addition) of two 1s atomic orbitals, whereas the $1s\sigma_u$ antibonding MO is formed from the out of-phase overlap (i.e., subtraction) of two 1s atomic orbitals. Similar considerations apply to overlap of p orbitals, although now these may form both σ and π bonding and antibonding MOs.

The stability of simple diatomic molecules such as H_2, H_2^- and He_2 depends on the relative filling of the bonding and antibonding MOs; e.g., H_2^- contains three electrons, two of which fill the bonding MO ($1s\sigma_g$ level) while the third enters the antibonding MO ($1s\sigma_u$ level); consequently, the overall bond strength is approximately half that in H_2. Meanwhile He_2 is unstable as there are an equal number of electrons in bonding MOs as in antibonding MOs. The same principles apply to more complicated diatomic molecules. However, if the atoms are different then the energy levels of the electrons associated with the constituent atoms will also be different and this will lead to an asymmetry in the MO energy level diagram.

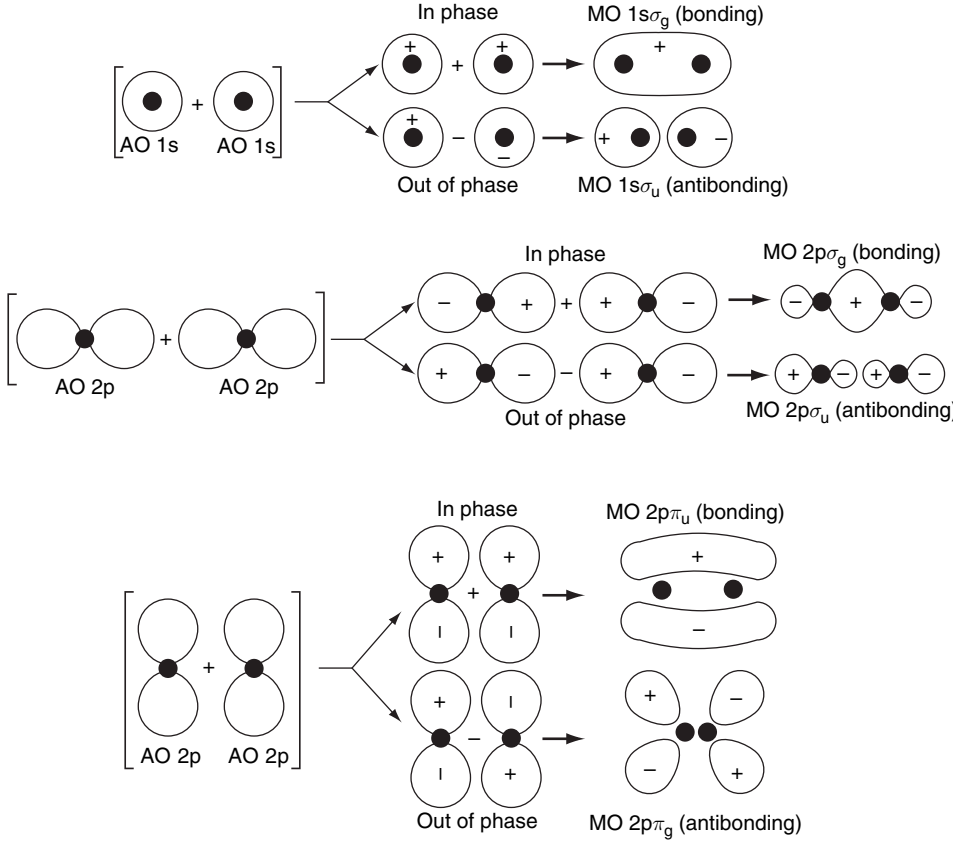


Figure 1.4 Formation of molecular orbitals from a linear combination of atomic orbitals; the + and − signs indicate the signs (phases) of the wavefunctions

For polyatomic molecules such as BF_3 a greater variety of molecular orbitals can be formed. MO theory emphasizes the delocalized nature of the electron distribution, so in general these MOs are extended over not just two, but all the constituent atoms. The total number of MOs (bonding, antibonding or non-bonding) is equal to the number of valence atomic orbitals used to construct them.

1.2.3 Giant molecular solids

When atoms come into close proximity with other atoms in a solid, most of the electrons remain localized and may be considered to remain associated with a particular atom. However, some outer electrons will become involved in bonding with neighbouring atoms. Upon bonding the atomic energy level diagram is modified. Briefly, the well-defined outer electron states of the atom overlap with those on neighbouring atoms and become broadened into energy bands. One convenient way of picturing this is to envisage the solid as a large molecule. Figure 1.5 shows the effect of increasing the number of atoms on the electronic energy levels of a one-dimensional solid (a linear chain of atoms).

For a simple diatomic molecule, as discussed previously, the two outermost atomic orbitals (AOs) overlap to produce two molecular orbitals (MOs) which can be viewed as a linear combination of the two constituent atomic orbitals. As before, the bonding MO is formed from the in-phase overlap of the AOs and is lower in energy than the corresponding AOs, whereas the other MO, formed from the out-of-phase overlap, is higher in energy than the corresponding AOs and is termed an antibonding MO. Progressively increasing the length of the molecular chain increases the total number of MOs, and gradually these overlap to form bands of allowed energy levels which are separated by forbidden energy regions (band gaps). These band gaps may be thought of as arising from the original energy gaps between the various atomic orbitals of the isolated atoms.

Note that the broadening of atomic orbitals into energy bands as the atoms are brought closer together to form a giant molecular solid can sometimes result in the overlapping of energy bands to give bands of mixed (atomic) character. The degree to which the orbitals are concentrated at a particular energy is reflected in a quantity known as the density of states (DOS) $N(E)$, where $N(E)\,dE$ is the number of allowed

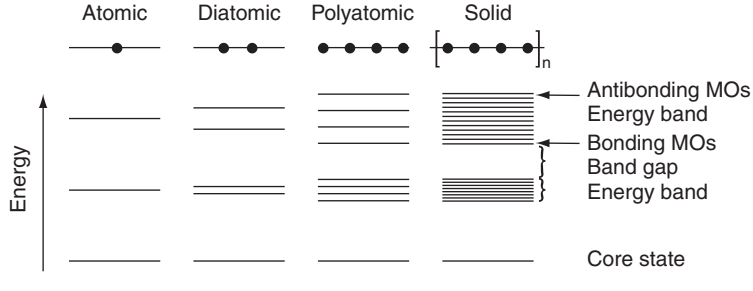


Figure 1.5 Electron energy level diagram for a progressively larger linear chain of atoms showing the broadening of molecular orbitals into energy bands for a one-dimensional solid

energy levels per unit volume of the solid in the energy range between E and $E + dE$. As in a simple molecule, each MO energy level in the energy band can accommodate two electrons of opposite spin. The total number of electrons from all the interacting atomic orbitals in the large molecule fill this set of MOs, the highest occupied energy level being known as the Fermi level E_F. The sum of the energies of all the individual electrons in the large molecule gives the total energy of the system, which gives a measure of the stability of the atomic arrangement in terms of the system free energy.

1.2.4 The free electron model and energy bands

An alternative view of the electronic band structure of solids is to consider the electron waves in a periodic crystalline potential. The starting point for this approach is the Drude–Lorentz free electron model for metals. In this model a metallic solid is considered as consisting of a close packed lattice of positive cations surrounded by an electron sea or cloud formed from the ionization of the outer shell (valence) electrons. We can then treat the valence electrons as if they were a gas inside a container and use classical kinetic gas theory. This works best for the electropositive metals of Groups I and II as well as aluminium (the so-called free electron metals) and can explain many of the fundamental properties of metals such as high electrical and thermal conductivities, optical opacity, reflectivity, ductility and alloying properties.

However, a more realistic approach is to treat the free electrons in metals quantum mechanically and consider their wave-like properties. Here the free valence electrons are assumed to be constrained within a potential well which essentially stops them from leaving the metal (the 'particle-in-a-box' model). The box boundary conditions require the wavefunctions to vanish at the edges of the crystal (or 'box'). The allowed wavefunctions given by the Schrödinger equation then correspond to certain wavelengths as shown in Figure 1.6. For a one-dimensional box of length L, the permitted wavelengths are $\lambda_n = 2L/n$, where $n = 1, 2, 3 \ldots$ is the quantum number of the state; the permitted wavevectors $k_n = 2\pi/\lambda$ are given by $k_n = n\pi/L$.

This simple particle-in-a-box model results in a set of wavefunctions given by

$$\psi_n = (2/L)^{1/2} \sin(n\pi x/L), \tag{1.4}$$

where $n = 1, 2, 3 \ldots .$, and for each n the corresponding energy of the electronic level is

$$E_n = \frac{n^2 h^2}{8mL^2}. \tag{1.5}$$

E_n represents solely kinetic energy since the potential energy is assumed to be zero within the box. Thus there is a parabolic relationship between E_n and n, and therefore between E_n and k since k depends directly on n as described above. The permitted energy levels on this parabola are discrete (i.e., quantized): however in principle the size of L for most metal crystals (ranging from microns to millimetres or even centimetres) means that the separation between levels is very small compared with the thermal energy $k_B T$,

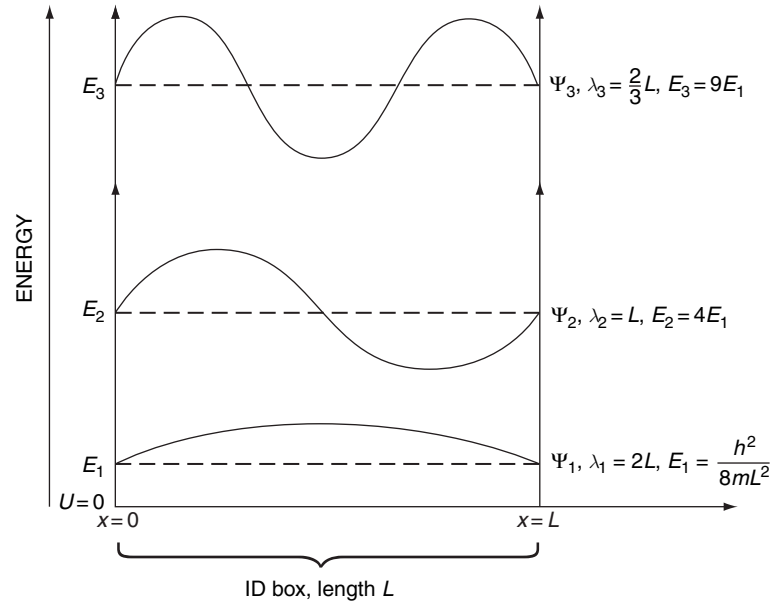


Figure 1.6 Energy level diagram also showing the form of some of the allowed wavefunctions for an electron confined to a one-dimensional potential well

and we can regard the energy distribution as almost continuous (quasi-continuous) so that the levels form a band of allowed energies as shown in figure 1.7.

Note that as the electron becomes more localized (i.e., L decreases), the energy of a particular electron state (and more importantly the spacing between energy states) increases; this has important implications for bonding and also for reduced-dimensionality or quantum-confined systems which are discussed later.

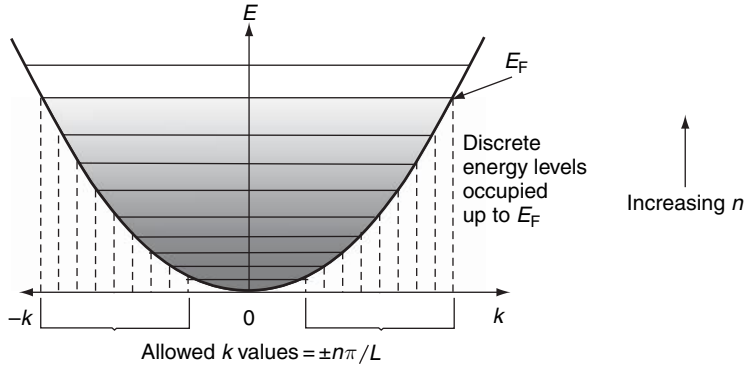


Figure 1.7 Schematic version of the parabolic relationship between the allowed electron wave vectors and the their energy for electrons confined to a one-dimensional potential well. Shaded energy regions represent those occupied with electrons

1.2.5 Crystalline solids

The above arguments may be extended from one to three dimensions to consider the electronic properties of bulk crystalline solids. For a perfectly ordered three dimensional crystal, the periodic repetition of atoms (or molecules) along the one dimensional linear chain considered in Section 1.2.2 is replaced by the periodic repetition of a *unit cell* in all three dimensions. The unit cell contains atoms arranged in the characteristic configuration of the crystal, such that contiguous replication of the unit cell throughout all space is sufficient to generate the entire crystal structure. In otherwords, the crystal has translational symmetry, and the crystal structure may be generated by translations of the unit cell in all three dimensions. Translation symmetry in a periodic structure is a so-called discrete symmetry, because only certain translations – those corresponding to integer multiples of the *lattice translation vectors* derived from the unit cell – lead to symmetry-equivalent points. (This may be contrasted with the case of empty space, which displays a continuous translation symmetry because *any* translation leads to a symmetry-equivalent point.) Common unit cells are simple cubic, face centred cubic, body centred cubic, and the diamond structure, which comprises two interlocking face-centred cubic lattices. However, in general, the lattice spacing may be different along the different principal axes, giving rise to the orthorhombic and tetragonal unit cells, and sides of the unit cell may not necessarily be orthogonal, such as in the hexagonal unit cell (refer to the Bibliography for further reading on this topic).

Generally, symmetries generate conservation laws; this is known as *Noether's theorem*. The continuous translation symmetry of empty space generates the law of momentum conservation; the weaker discrete translation symmetry in crystals leads to a weaker quasi-conservation law for quasi- or crystal momentum. An important consequence of discrete translation symmetry for the electronic properties of crystals is *Bloch's theorem*, which is described below.

1.2.6 Periodicity of crystal lattices

The three dimensional periodicity of the atomic arrangement in a crystal gives rise to a corresponding periodicity in the internal electric potential due to the ionic cores. Incorporating this periodic potential into the Schrödinger equation results in allowed wavefunctions that are modulated by the lattice periodicity. Bloch's theorem states that these wavefunctions take the form of a plane wave (given by $\exp(\mathrm{i}\mathbf{k}.\mathbf{r})$) multiplied by a function which has the same periodicity as the lattice; i.e.,

$$\psi(\mathbf{r}) = u_k(\mathbf{r}) \exp(\mathrm{i}\mathbf{k}.\mathbf{r}), \tag{1.6}$$

where the function $u_k(\mathbf{r})$ has the property $u_k(\mathbf{r}+\mathbf{T}) = u_k(\mathbf{r})$, for any lattice translation vector $\mathbf{T}$. Such wavefunctions are known as Bloch functions, and represent travelling waves passing through the crystal, but with a form modified periodically by the crystal potential due to each atomic site. For a one dimensional lattice of interatomic spacing a, these relationships reduce to

$$\psi(x) = u_k(x) \exp(\mathrm{i}kx) \tag{1.7}$$

with $u_k(x+na)=u_k(x)$ for any integer n. Now, if we impose periodic boundary conditions at the ends of the chain of atoms of length $L=Na$:

$$\psi(Na)=\psi(0), \tag{1.8}$$

we find that

$$u_k(Na)\exp(\mathrm{i}kNa)=u_k(0), \tag{1.9}$$

from which $\exp(\mathrm{i}kNa)=1$, which has the solutions

$$k=\pm 2n\pi/Na=\pm 2n\pi/L \tag{1.10}$$

for integer n. This result has two important consequences: firstly, it tells us that the difference between consecutive k values is always $2\pi/L$, which can be interpreted as representing the volume (or properly, in this simplified one dimensional case, the *length*) of **k**-space occupied by each wavevector state. Applying this argument to each dimension in turn gives, for a 3D crystal, a **k**-space volume of $8\pi^3/V$ occupied by each wavevector state, where $V=L^3$ is the crystal volume. Secondly, once the upper limit on n is determined, equation (1.10) will also tell us how many wavevector states are contained within each energy band. This point is examined below.

The lattice periodicity also gives rise to diffraction effects. Diffraction of X-rays in crystals is discussed in detail in Chapter 2, as an important structural characterisation technique. However, since electrons exhibit wave-like properties, the free electrons present in the crystal also experience the same diffraction phenomena, and this has a crucial effect on the spectrum of allowed electron energies. If we consider an electron wave travelling along a one dimensional chain of atoms of spacing a, then each atom will cause reflection of the wave. These reflections will all be constructive provided that $m\lambda=2a$, for integer m, where λ is the electron de Broglie wavelength (this is a special case of the Bragg Law of diffraction introduced in Section 2.1.2.5). When this condition is satisfied, both forward and backward travelling waves exist in the lattice, and the superposition of these creates standing waves. The standing waves correspond to electron density distributions $|\psi(x)|^2$ which have either all nodes, or all antinodes, at the lattice sites $x=a,\ 2a,\ 3a,\ldots$, and these two solutions, although having the same wavevector value, have quite different associated energies, due to the different interaction energies between the electrons and the positively charged ions. Consequently a *band gap* forms in the electron dispersion curve at the corresponding values of wavevector: $k=\pm m\pi/a$ (see figure 1.8). The fact that the electron waves are standing waves means that the electron group velocity

$$\nu_g=\frac{\partial\omega}{\partial k}=\frac{1}{\hbar}\frac{\partial E}{\partial k} \tag{1.11}$$

tends to zero at these points. This represents a fundamental difference between the behaviour of electrons in crystalline solids and that in free space, where the dispersion relationship remains purely parabolic ($E\propto k^2$) for all values of k.

The region of **k** space which lies between any two diffraction conditions is know as a Brillouin zone: thus, in a one-dimensional crystal, the first Brillouin zone lies between $k=-\pi/a$ and $k=+\pi/a$. However, any value of k which lies outside the first Brillouin zone corresponds, mathematically, to an electron wave of wavelength $\lambda<2a$. Such

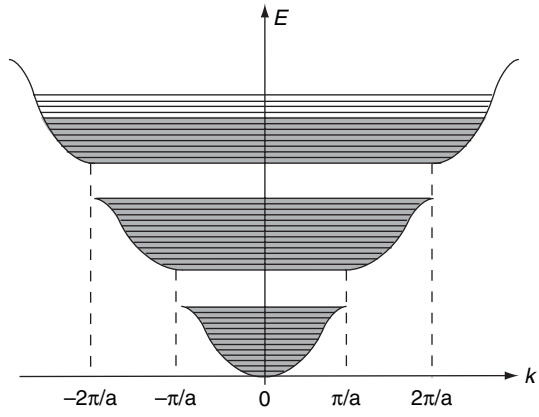


Figure 1.8 Schematic version of the parabolic relationship between the allowed electron wavevectors and the their energy for electrons confined to a one-dimensional potential well containing a periodically varying potential of period a. Shaded energy regions represent those occupied with electrons

a wave has too high a spatial frequency to be uniquely defined by a set of wave amplitudes which are only specified at lattice sites: an equivalent wave of wavelength $\lambda > 2a$ can always be identified. In **k**-space, this transformation is represented by the fact that any value of k lying outside the first Brillouin zone is equivalent to some point lying inside the first Brillouin zone, where the equivalent point is found from the relation

$$k' = k \pm 2m\pi/a \tag{1.12}$$

and the set of values $2m\pi/a$ are known as the reciprocal lattice vectors for the crystal.

In a three dimensional crystal, the location of energy gaps in the electron dispersion is still determined from electron diffraction by the lattice planes, but the Brillouin zones are no longer simple ranges of k, as in 1D: rather, they are described by complex surfaces in 3D **k**-space, the geometry of which depends on the unit cell and atomic structure. When the energy–wavevector relationship for such a crystal extending over multiple Brillouin zones is mapped entirely into the first Brillouin zone, as described above, this results in a large number of different energy bands and consequently the density of energy states takes on a very complex form. An example of the multiple energy bands and corresponding density of states in a real crystal is shown for the case of silicon in figure 1.9.

1.2.7 Electronic conduction

We may now observe that the series of allowed k values in equation 1.12 extends up to the edges of the Brillouin zone, at $k = \pm\pi/a$. Since one of these endpoints may be mapped onto the other by a reciprocal lattice vector translation, the total number of allowed k values is precisely N. Recalling that each k state may be occupied by both a spin up and a spin down electron, the total number of states available is $2N$ per energy band. In three dimensions, this result is generalised to $2N_u$ states per band, where N_u is the number of unit cells in the crystal. Now, the total number of valence electrons in the

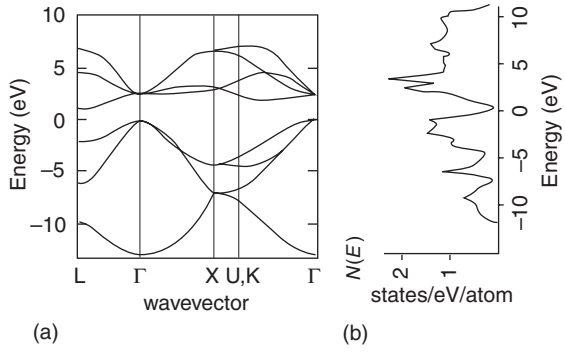


Figure 1.9 Electron energy band structure diagram and density of states for crystalline silicon. The symbols Γ, L, X, U and K on the horizontal (wavevector) axis of the band structure plot represent different symmetry points in 3D $\boldsymbol{k}$-space. Γ corresponds to $\boldsymbol{k} = \mathbf{0}$, the origin of the Brillouin zone; the range Γ–X represents a path through the Brillouin zone from centre to edge along the 100 direction; Γ–L and Γ–K represent middle to edge paths along the ⟨111⟩ and ⟨110⟩ directions, respectively, and X–U represents a path along the Brillouin zone boundary starting from the zone edge on the ⟨100⟩ axis and moving in a direction parallel to ⟨101⟩

crystal is zN_u, where z is the number of valence electrons per unit cell. This leads to two very different electronic configurations in a solid. If z is even, then one energy band is completely filled, with the next band being completely empty. The highest filled band is the valence band, and the next, empty band, is the conduction band. The electrons in the valence band cannot participate in electrical conduction, because there are no available states for them to move into consistent with the small increase in energy required by motion in response to an externally applied voltage: hence this configuration results in an insulator or, if the band gap is sufficiently small, a semiconductor. Alternatively, if z is odd, then the highest occupied energy band is only half full. In such a material, there are many vacant states immediately adjacent in energy to the highest occupied states, therefore electrical conduction occurs very efficiently and the material is a metal. Figure 1.10 shows schematic energy diagrams for insulators, metals and semiconductors respectively. There is one further, special case which gives rise to metallic behaviour: namely, when the valence band is completely full (z is even), but the valence and

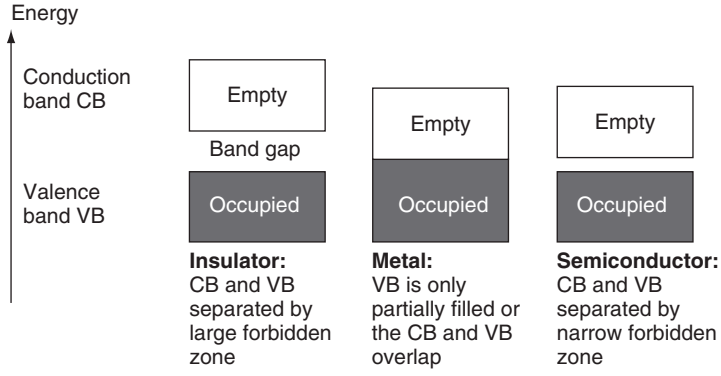


Figure 1.10 Electron energy band diagram for an insulator, a conductor and a semiconductor

conduction bands overlap in energy, such that there are vacant states immediately adjacent to the top of the valence band, just as in the case of a half-filled band. Such a material is called a semi-metal.

In the same way as was defined in the molecular orbital theory of section 1.2.3, the uppermost occupied energy level in a solid is the Fermi level E_F, and the corresponding Fermi wavevector is given by $E_F = \hbar k_F^2/2m_e$. As mentioned above, the volume of **k**-space per state is $8\pi^3/V$. Therefore, the volume of **k** space filled by N electrons is $4N\pi^3/V$, accounting for the fact that 2 electrons of opposite spins can occupy each wavevector state. If we equate this volume to the volume of a sphere in **k** space, of radius k_F (the Fermi sphere), we obtain the result

$$k_F = (3\pi^2 n_e)^{1/3} \tag{1.13}$$

where $n_e = N/V$ is the electron density, and hence

$$E_F = \hbar(3\pi^2 n_e)^{2/3}/2m_e. \tag{1.14}$$

If the Fermi sphere extends beyond the first Brillouin zone, as occurs in many metals, then the appropriate mapping back into the zone results in a Fermi surface of complex topology.

The density of states $N(E)dE$ is defined such that $N_s = \int N(E)dE$ gives the total number of states per unit crystal volume in an energy band. Now, from the above argument, the number of wavevector states per unit volume of **k** space is $V/8\pi^3$. Thus, the total number of states per band may be calculated from

$$N_s = 2 \times V/8\pi^3 \int d\mathbf{k} \tag{1.15}$$

where the factor of 2 accounts for the 2 spin states per **k** value. In three dimensions, $d\mathbf{k} = 4\pi k^2 dk$ and thus we may write

$$N_s = \frac{V}{4\pi^3} \int 4\pi k^2 \frac{dk}{dE} dE. \tag{1.16}$$

For parabolic bands, $E = \hbar^2 k^2/(2m_e)$ and hence $dk/dE = m_e/(\hbar^2 k)$, from which

$$N(E) = \frac{4\pi(2m_e)^{3/2}E^{1/2}}{h^3}. \tag{1.17}$$

The dependence of the density of states on $E^{1/2}(\propto k)$ is simply a consequence of the increased volume of phase space available at larger values of energy. The actual population of electrons as a function of energy is given by the product of the density of states and the occupation probability $f(E)$ which, for electrons or holes, is given by the Fermi Dirac function

$$f(E) = \frac{1}{\exp((E - E_F)/k_B T) + 1}, \tag{1.18}$$

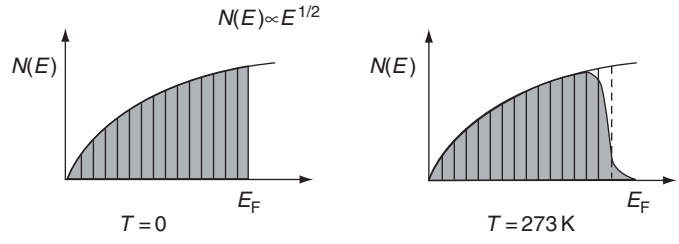


Figure 1.11 The density of electron states for free electrons and the occupation of electron energy levels (shaded region) at zero and room temperature

from which we may observe that the Fermi energy corresponds to the energy at which the occupation probability is exactly one half.

In the zero temperature limit, $f(E) = 1$ for all $E < E_F$, and $f(E) = 0$ for all $E > E_F$: in otherwords, all electron states below the Fermi energy are filled, and all those above E_F are empty, as previously described; the electron population at any energy $E < E_F$ is then given just by $N(E)$. At non-zero temperatures $f(E)$ describes the fact that some electrons are thermally excited from states just below E_F to states just above E_F, and the sharpness of the cut-off of $N(E)$ at E_F decreases with increasing temperature. Both zero temperature and non-zero temperature cases are shown in Figure 1.11.

In addition to the total DOS, which has already been mentioned, it is possible to project the DOS onto a particular atomic site in the unit cell and determine the so-called local DOS; this is the contribution of that particular atomic site to the overall electronic structure. If a unit cell contains a particular type of atom in two distinct crystallographic environments, then the local DOS will be correspondingly different. Similar projections may be performed in terms of the angular momentum symmetry (i.e., the s, p, d or f atomic character of the DOS).

Until now we have been concerned with crystalline systems. However, it is also possible to consider the DOS of an amorphous material; here the DOS is primarily determined by the short-range order in the material; i.e., the nearest neighbours. An alternative approach is to represent the amorphous solid by a very large unit cell with a large number of slightly different atomic environments.

1.3 EFFECTS OF THE NANOMETRE LENGTH SCALE

The small length scales present in nanoscale systems directly influence the energy band structure and can lead indirectly to changes in the associated atomic structure. Such effects are generally termed quantum confinement. The specific effects of quantum confinement in one, two and three dimensions on the density of states are discussed in detail in the Chapter 3 for the case of semiconductor nanostructures; however, initially we outline two general descriptions that can account for such size-dependent effects in nanoscale systems.

1.3.1 Changes to the system total energy

In the free electron model, it is clear that the energies of the electronic states depend on $1/L^2$ where L is the dimension of the system in that particular direction; the spacing between successive energy levels also varies as $1/L^2$. This behaviour is also clear from the description of a solid as a giant molecule: as the number of atoms in the molecule increases, the MOs gradually move closer together. Thus if the number of atoms in a system, hence the length scale, is substantially different to that in a normal bulk material, the energies and energy separations of the individual electronic states will be very different. Although in principle the Fermi level (Section 1.2.5) would not be expected to change since the free electron density N/V should remain constant, there may be associated modifications in structure (see below) which will change this quantity. Furthermore, as the system size decreases, the allowed energy bands become substantially narrower than in an infinite solid. The normal collective (i.e., delocalized) electronic properties of a solid become severely distorted and the electrons in a reduced-dimensional system tend to behave more like the 'particle in a box' description (Section 1.2.5); this is the phenomenon of quantum confinement. In otherwords, the electronic states are more like those found in localized molecular bonds rather than those in a macroscopic solid.

The main effect of these alterations to the bulk electronic structure is to change the total energy and hence, ignoring entropy considerations, the thermodynamic stability of the reduced length scale system relative to that of a normal bulk crystal. This can have a number of important implications. It may change the most energetically stable form of a particular material; for example, small nanoparticles or nanodimensional layers may adopt a different crystal structure from that of the normal bulk material. For example, some metals which normally adopt a hexagonal close-packed atomic arrangement have been reported to adopt a face-centred cubic structure in confined systems such as metallic multilayers. If a different crystallographic structure is adopted below some particular critical length scale, then this arises from the corresponding change in the electronic density of states which often results in a reduced total energy for the system.

Reduction of system size may change the chemical reactivity, which will be a function of the structure and occupation of the outermost electronic energy levels. Correspondingly, physical properties such as electrical, thermal, optical and magnetic characteristics, which also depend on the arrangement of the outermost electronic energy levels, may be changed. For example, metallic systems can undergo metal–insulator transitions as the system size decreases, resulting from the formation of a forbidden energy band gap. Other properties such as mechanical strength which, to a first approximation, depends on the change in electronic structure as a function of applied stress and hence interatomic spacing, may also be affected. Transport properties may also change in that they may now exhibit a quantized rather than continuous behaviour, owing to the changing nature and separation of the electron energy levels.

1.3.2 Changes to the system structure

A related viewpoint for understanding the changes observed in systems of reduced dimension is to consider the proportion of atoms which are in contact with either a

free surface, as in the case of an isolated nanoparticle, or an internal interface, such as a grain boundary in a nanocrystalline solid. Both the surface area to volume ratio (S/V) and the specific surface area (m^2g^{-1}) of a system are inversely proportional to particle size and both increase drastically for particles less than 100 nm in diameter. For isolated spherical particles of radius r and density ρ, the surface area per unit mass of material is equal to $4\pi r^2/(4/3\pi r^3\rho) = 3/r\rho$. For 2 nm diameter spherical particles of typical densities, the specific surface area (SSA) can approach $500\,m^2\,g^{-1}$. However, for particles in contact this value will be reduced by up to approximately half. This large surface area term will have important implications for the total energy of the system. As discussed above this may lead to the stabilization of metastable structures in nanometre-sized systems, which are different from the normal bulk structure or, alternatively, may induce a simple relaxation (expansion or contraction) of the normal crystalline lattice which could in turn alter other material properties.

If an atom is located at a surface then it is clear that the number of nearest-neighbour atoms are reduced, giving rise to differences in bonding (leading to the well-known phenomenon of surface tension or surface energy) and electronic structure. In a small isolated nanoparticle, a large proportion of the total number of atoms will be present either at or near the free surface. For instance, in a 5 nm particle approximately 30–50% of the atoms are influenced by the surface, compared with approximately a few percent for a 100 nm particle. Similar arguments apply to nanocrystalline materials, where a large proportion of atoms will be either at or near grain boundaries. Such structural differences in reduced-dimensional systems would be expected to lead to very different properties from the bulk.

1.3.2.1 Vacancies in nanocrystals

Another important consideration for nanostructures concerns the number of atomic vacancies n_v which exist in thermal equilibrium in a nanostructure. Vacancies are point defects in the crystalline structure of a solid and may control many physical properties in materials such as conductivity and reactivity. In microcrystalline solids at temperatures above 0 K, vacancies invariably exist in thermal equilibrium. In the simple case of metals with one type of vacancy, the number of vacancies in a crystal consisting of N atom sites is approximated by an Arrhenius-type expression

$$n_v = N\exp(-Q_f/RT), \tag{1.19}$$

where T is the absolute temperature, R is the gas constant and Q_f is the energy required to form one mole of vacancies. Q_f is given by the relationship $Q_f = N_A q_f$, where N_A is the Avogadro number and q_f is the activation energy for the formation of one vacancy. However, the value of q_f is not well defined but is generally estimated to be the energy required to remove an atom from the bulk interior of a crystal to its surface. As a rough approximation, a surface atom is bonded to half the number of atoms compared with an interior atom, so q_f represents half the bonding energy per atom. Since the melting temperature T_m of a metal is also a measure of the bond energy, then q_f is expected to be a near linear function of T_m.

From a continuum model, Q_f may be estimated from the latent heat of vaporisation, since on leaving the surface an atom breaks the remaining (half) bonds. In practice it is found that the latent heat of vaporisation is considerably higher than Q_f. Alternatively, Q_f may be estimated from the surface energy per unit area. Given that one atom occupies an area b^2, the number of atoms per unit area is equal to $1/b^2$ and the surface energy σ is therefore q_f/b^2. Surface energies depend on melting temperature and vary within the range $1.1\,J\,m^{-2}$ (for aluminium) to $2.8\,J\,m^{-2}$ (for tungsten). Taking an average value of σ as $2.2\,J\,m^{-2}$ and $b = 2.5 \times 10^{-10}$ m, we may calculate $Q_f = N_A \sigma b^2$ as $83 \times 10^3\,J\,mol^{-1}$, which is close to the accepted value of $90\,kJ\,mol^{-1}$.

Furthermore, the value of Q_f may be modified for nanoparticles through the influence of the surface energy term, σ, which is related to the internal pressure, P, by the simple relationship $P = 4\sigma/d$, where d is the diameter of the nanoparticle. The effect of P is to require an additional energy term, q_n, for the formation of a vacancy, which is approximately given by Pb^3. Again taking σ as $2.2\,J\,m^{-2}$, we may calculate this additional energy per mole $Q_n = N_A q_n$ as $8.3 \times 10^3\,J\,mol^{-1}$ for a 10 nm diameter nanoparticle. This term is only approximately 10% of Q_f and rapidly decreases for larger particle sizes. Thus we may conclude that the effect of the surface energy (internal pressure) factor on the vacancy concentration will be small. Additionally, the internal pressure P results in an elastic, compressive volume strain, and hence linear strain, ε, given approximately as

$$\varepsilon = P/3E = 4\sigma/3dE \tag{1.20}$$

where E is the Young's modulus. This expression suggests that the linear strain will be inversely proportional to particle size and that there will be a decrease in lattice parameter or interatomic spacing for small nanoparticles. This prediction correlates reasonably well with the data in Figure 1.12.

Finally, substituting a value of $Q_f = 90 \times 10^3\,J\,mol^{-1}$ into the Arrhenius expression (1.19) for the vacancy concentration, we obtain values for the ratio n_v/N of 2.4×10^{-16} (at 300 K), 6.5×10^{-7} (at 600 K) and 4.8×10^{-4} (at 1000 K), illustrating the exponential

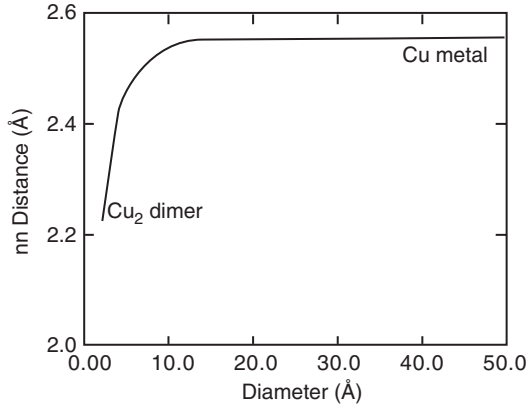


Figure 1.12 Schematic diagram of the change in nearest-neighbour (nn) distance as a function of cluster size or particle size for copper

increase in vacancy concentration with temperature. Now consider a spherical nanoparticle, say 50 nm in diameter, and in which each atom occupies a volume b^3. Again taking $b = 0.25$ nm, there are a total of 4.2×10^6 atoms in the particle, which implies that $n_v << 1$, except at very high temperatures. Therefore nanocrystals are predicted to be essentially vacancy-free; their small size precludes any significant vacancy concentration. This simple result also has important consequences for all thermomechanical properties and processes (such as creep and precipitation) which are based on the presence and migration of vacancies in the lattice.

1.3.2.2 Dislocations in nanocrystals

Planar defects, such as dislocations, in the crystalline structure of a solid are extremely important in determining the mechanical properties of a material. It is expected that dislocations would have a less dominant role to play in the description of the properties of nanocrystals than in the description of the properties of microcrystals, owing to the dominance of crystal surfaces and interfaces. The free energy of a dislocation is made up of a number of terms: (i) the core energy (within a radius of about three lattice planes from the dislocation core); (ii) the elastic strain energy outside the core and extending to the boundaries of the crystal, and (iii) the free energy arising from the entropy contributions. In microcrystals the first and second terms increase the free energy and are by far the most dominant terms. Hence dislocations, unlike vacancies, do not exist in thermal equilibrium.

The core energy is expected to be independent of grain size. Estimates are close to 1 eV per lattice plane which, for an interplanar spacing b of 0.25 nm, translates to a value of about 6.5×10^{-10} J m^{-1}. The elastic strain energy per unit length for an edge dislocation is given by

$$E = \frac{Gb^2}{4\pi(1-\nu)} \times \ln\left(\frac{r_1}{r_0}\right) \tag{1.21}$$

where G is the bulk modulus, r_0 is the core radius, r_1 is the crystal radius and ν is Poisson's ratio. ν is typically around 1/3 for a crystalline sample. The expression for a screw dislocation omits the $(1-\nu)$ term, giving an energy about 2/3 that of an edge dislocation. For $G = 40 \times 10^9$ Pa, the constant term $Gb^2/4\pi(1-\nu)$ has a value of 3×10^{-10} J m^{-1}. The grain size dependence is given in the $\ln(r_1/r_0)$ term, which for grain size $(2r_1)$ values of 10, 50, 1000 and 10 000 nm increases as 3, 4.6, 7.6 and 9.9 respectively. Hence it can be seen that the elastic strain energy of dislocations in nanoparticles and nanograined materials is about one-third of that in microcrystals and that, for a 10 nm grain size the core energy is comparable with the elastic strain energy. In comparison, the core energy is about one-tenth of the elastic strain energy for a microcrystal.

This reduction in the elastic strain energy of dislocations in nanocrystals has important consequences. The forces on dislocations due to externally applied stresses are reduced by a factor of about three and the interactive forces between dislocations are reduced by a factor of about 10. Hence recovery rates and the annealing out of dislocations to free surfaces are expected to be reduced. For a dislocation near the surface of a semi-infinite solid, the stress towards the surface is given by the interaction

of the stress field of an image dislocation at an equal distance on the opposite side. Since nanocrystals do not approximate to semi-infinite solids, such image stresses will operate across all surfaces and the net effect, together with the reduced elastic strain energy, results in dislocations that are relatively immobile.

Finally, we estimate the entropy contributions to the free energy. These arise as a result of (i) configurational entropy (i.e., the dislocation can be arranged in a variety of ways), (ii) a further contribution if the dislocation is assumed to be perfectly flexible, and (iii) the effect of the dislocation on the thermal vibrations of the crystal. Factors (ii) and (iii) are independent of crystal size and their values can be estimated to be $2k_BT$ and $3k_BT$, respectively, per atomic plane. Assuming a temperature of 300 K, these values correspond to about $3 \times 10^{-11}\,\mathrm{J\,m^{-1}}$, considerably less than the core and elastic strain energy terms. The configurational entropy contribution to the free energy is given by

$$E = \frac{bk_BT}{L}\ln\left(\frac{L^2}{b^2}\right) \tag{1.22}$$

per atom plane, where L is the length of the dislocation. At 300 K this gives values of $3.0 \times 10^{-12}, 5.7 \times 10^{-14}$ and $7.6 \times 10^{-15}\,\mathrm{J\,m^{-1}}$ for $L = 10$, 1000 and 10 000 nm, respectively. These values are again much smaller than the core and elastic strain energy terms and hence it may be concluded that dislocations in nanocrystals, as with microcrystals, do not exist as thermodynamically stable lattice defects.

1.3.3 How nanoscale dimensions affect properties

Many properties are continuously modified as a function of system size. Often these are extrinsic properties, such as resistance, which depend on the exact size and shape of the specimen. Other properties depend critically on the microstructure of the material; for example, the Hall–Petch equation for yield strength, σ, of a material as a function of average grain size $\langle d \rangle$ is given by

$$\sigma = k\langle d \rangle^{-1/2} + \sigma_0 \tag{1.23}$$

where k and σ_0 are constants. Intrinsic materials properties, such as resistivity, should be independent of specimen size, however, even many of the intrinsic properties of matter at the nanoscale are not necessarily predictable from those observed at larger scales. As discussed above this is because totally new phenomena can emerge, such as: quantum size confinement leading to changes in electronic structure; the presence of wave-like transport processes, and the predominance of interfacial effects.

1.3.3.1 Structural properties

The increase in surface area and surface free energy with decreasing particle size leads to changes in interatomic spacings. For Cu metallic clusters the interatomic spacing is observed to decrease with decreasing cluster size, as shown in Figure 1.12. This effect

can be explained by the compressive strain induced by the internal pressure arising from the small radius of curvature in the nanoparticle (Section 1.3.2.2). Conversely, for semiconductors and metal oxides there is evidence that interatomic spacings increase with decreasing particle size.

A further effect, previously mentioned, is the apparent stability of metastable structures in small nanoparticles and clusters, such that all traces of the usual bulk atomic arrangement become lost. Metallic nanoparticles, such as gold, are known to adopt polyhedral shapes such as cube–octahedra, multiply twinned icosahedra and multiply twinned decahedra (Figure 1.13). These nanoparticles may be regarded as multiply twinned crystalline particles (MTPs) in which the shapes can be understood in terms of the surface energies of various crystallographic planes, the growth rates along various crystallographic directions and the energy required for the formation of defects such as twin boundaries. However, there is compelling evidence that such particles are not crystals but are quasiperiodic crystals or crystalloids. These icosahedral and decahedral quasicrystals form the basis for further growth of the nanocluster, up until a size where they will switch into more regular crystalline packing arrangements.

Crystalline solids are distinct from amorphous solids in that they possess long-range periodic order and the patterns and symmetries which occur correspond to those of the 230 space groups. Quasiperiodic crystals do not possess such long-range periodic order and are distinct in that they exhibit fivefold symmetry, which is forbidden in the 230 space groups. In the cubic close-packed and hexagonal close-packed structures, exhibited by many metals, each atom is coordinated by 12 neighbouring atoms. All of the coordinating atoms are in contact, although not evenly distributed around the central atom. However, there is an alternative arrangement in which each coordinating atom is situated at the apex of an icosahedron and in contact only with the central atom. If however we relax this 'rigid atomic sphere' model and allow the central atom to reduce in diameter by 10%, the coordinating atoms come into contact and the body now has the shape and symmetry of a regular icosahedron with point group symmetry 235, indicating the presence of 30 twofold, 20 threefold and 12 fivefold axes of symmetry. This geometry represents the nucleus of a quasiperiodic crystal which may grow in the forms of icosahedra or pentagonal dodecahedra. These are dual solids with identical symmetry, the apices of one being replaced by the faces of the other. Such quasiperiodic crystals are known to exist in an increasing number of aluminium-based alloys and may be stable up to microcrystalline sizes. It should be noted that their symmetry is precisely the same as that of the fullerenes C_{20} (dodecahedrene with 12 pentagonal faces of a pentagonal dodecahedron, but unstable) and C_{60} (the well-known buckyball with 12 pentagonal faces and 20 hexagonal faces of a truncated icosahedron). Hence, like the fullerenes, quasiperiodic crystals are expected to have an important role to play in nanostructures.

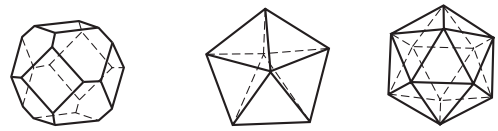


Figure 1.13 Geometrical shapes of cubo-octahedral particles and multiply twinned decahedral and icosahedral particles

The size-related instability characteristics of quasiperiodic crystals are not well understood. A frequently observed process appears to be that of multiple twinning, such crystals being distinguished from quasiperiodic crystals by their electron diffraction patterns. Here the five triangular faces of the fivefold symmetric icosahedron can be mimicked by five twin-related tetrahedra (with a close-packed crystalline structure) through relatively small atomic movements.

1.3.3.2 Thermal properties

The large increase in surface energy and the change in interatomic spacing as a function of nanoparticle size mentioned above have a marked effect on material properties. For instance, the melting point of gold particles, which is really a bulk thermodynamic characteristic, has been observed to decrease rapidly[2] for particle sizes less than 10 nm, as shown in Figure 1.14. There is evidence that for metallic nanocrystals embedded in a continuous matrix the opposite behaviour is true; i.e., smaller particles have higher melting points.[3]

1.3.3.3 Chemical properties

The change in structure as a function of particle size is intrinsically linked to the changes in electronic properties. The ionization potential (the energy required to remove an electron) is generally higher for small atomic clusters than for the corresponding bulk material. Furthermore, the ionization potential exhibits marked fluctuations as a function of cluster size. Such effects appear to be linked to chemical reactivity, such as the reaction of Fe_n clusters with hydrogen gas (Figure 1.15).

Nanoscale structures such as nanoparticles and nanolayers have very high surface area to volume ratios and potentially different crystallographic structures which

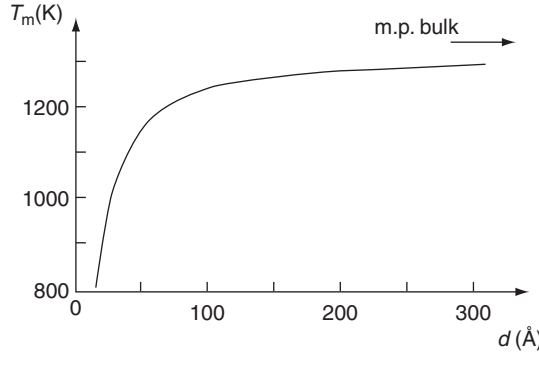


Figure 1.14 Schematic diagram of the variation in melting point of gold nanoparticles as a function of particle size

[2] *Nanomaterials: Synthesis, Properties and Applications*, ed. A. S. Edelstein and R. C. Cammarata (Institute of Physics 1996) and references therein.
[3] U. Dahmen *et al.*, *Inst.Phys. Conf. Ser.* **168**, 1 (IOP Publishing 2001).

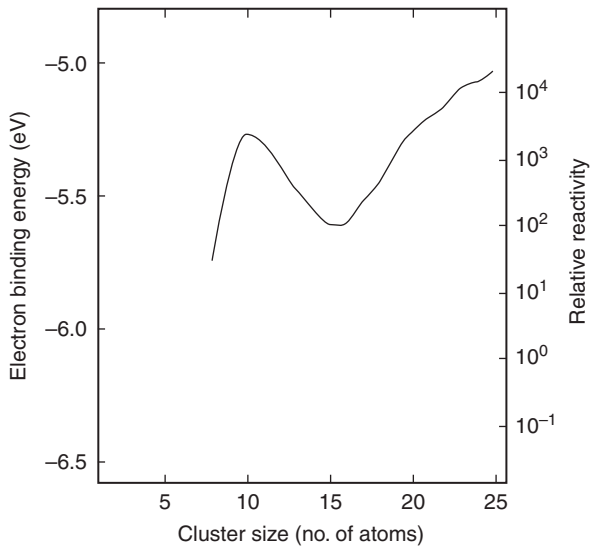


Figure 1.15 Schematic diagram of the dependence of the electron binding energy and relative chemical reactivity of iron clusters to hydrogen gas as a function of cluster size

may lead to a radical alteration in chemical reactivity. Catalysis using finely divided nanoscale systems can increase the rate, selectivity and efficiency of chemical reactions such as combustion or synthesis whilst simultaneously significantly reducing waste and pollution. Gold nanoparticles smaller than about 5 nm in diameter are known to adopt icosahedral structures rather than the normal face centred cubic arrangement. This structural change is accompanied by an extraordinary increase in catalytic activity. Furthermore, nanoscale catalytic supports with controlled pore sizes can select the products and reactants of chemical reactions based on their physical size and thus ease of transport to and from internal reaction sites within the nanoporous structure. Additionally, nanoparticles often exhibit new chemistries as distinct from their larger particulate counterparts; for example, many new medicines are insoluble in water when in the form of micron-sized particles but will dissolve easily when in a nanostructured form.

1.3.3.4 Mechanical properties

Many mechanical properties, such as toughness, are highly dependent on the ease of formation or the presence of defects within a material. As the system size decreases, the ability to support such defects becomes increasingly more difficult and mechanical properties will be altered accordingly. Novel nanostructures, which are very different from bulk structures in terms of the atomic structural arrangement, will obviously show very different mechanical properties. For example, single- and multi-walled carbon nanotubes show high mechanical strengths and high elastic limits that lead to considerable mechanical flexibility and reversible deformation.

As the structural scale reduces to the nanometre range, for example, in nanoayered composites, a different scale dependence from the usual Hall–Petch relationship

(Equation 1.23) for yield strength often becomes apparent with large increases in strength reported. In addition, the high interface to volume ratio of consolidated nanostructured materials appears to enhance interface-driven processes such as plasticity, ductility and strain to failure. Many nanostructured metals and ceramics are observed to be superplastic, in that they are able to undergo extensive deformation without necking or fracture. This is presumed to arise from grain boundary diffusion and sliding, which becomes increasingly significant in a fine-grained material. Overall these effects extend the current strength–ductility limit of conventional materials, where usually a gain in strength is offset by a corresponding loss in ductility.

1.3.3.5 Magnetic properties

Magnetic nanoparticles are used in a range of applications, including ferrofluids, colour imaging, bioprocessing, refrigeration as well as high storage density magnetic memory media. The large surface area to volume ratio results in a substantial proportion of atoms (those at the surface which have a different local environment) having a different magnetic coupling with neighbouring atoms, leading to differing magnetic properties. Figure 1.16 shows the magnetic moments of nickel nanoparticles as a function of cluster size.

Whilst bulk ferromagnetic materials usually form multiple magnetic domains, small magnetic nanoparticles often consist of only one domain and exhibit a phenomenon known as superparamagnetism. In this case the overall magnetic coercivity (Section 4.1) is then lowered: the magnetizations of the various particles are randomly distributed due to thermal fluctuations and only become aligned in the presence of an applied magnetic field.

Giant magnetoresistance (GMR) is a phenomenon observed in nanoscale multi-layers consisting of a strong ferromagnet (e.g., Fe, Co) and a weaker magnetic or non-magnetic buffer (e.g., Cr, Cu); it is usually employed in data storage and sensing. In the absence of a magnetic field the spins in alternating layers are oppositely aligned through antiferromagnetic coupling, which gives maximum scattering from the interlayer interface and hence a high resistance parallel to the layers. In an

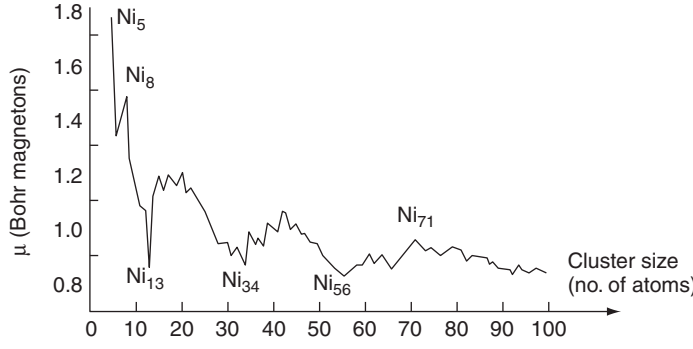


Figure 1.16 Schematic diagram of the variation in magnetic moments of clusters as a function of cluster size. The Bohr magneton is the classical magnetic moment associated with an electron orbiting a nucleus which has a single positive charge

oriented external magnetic field the spins align with each other and this decreases scattering at the interface and hence resistance of the device. Further details are given in Chapter 4.

1.3.3.6 Optical properties

In small nanoclusters the effect of reduced dimensionality on electronic structure has the most profound effect on the energies of the highest occupied molecular orbital (HOMO), essentially the valence band, and the lowest unoccupied molecular orbital (LUMO), essentially the conduction band. Optical emission and absorption depend on transitions between these states; semiconductors and metals, in particular, show large changes in optical properties, such as colour, as a function of particle size. Colloidal solutions of gold nanoparticles have a deep red colour which becomes progressively more yellow as the particle size increases; indeed gold colloids have been used as a pigment for stained glass since the seventeenth century. Figure 1.17 shows optical absorption spectra for colloidal gold nanoparticles of varying sizes. Semiconductor nanocrystals in the form of quantum dots show similar size-dependent behaviour in the frequency and intensity of light emission as well as modified non-linear optical properties and enhanced gain for certain emission energies or wavelengths. Other properties which may be affected by reduced dimensionality include photocatalysis, photoconductivity, photoemission and electroluminescence.

1.3.3.7 Electronic properties

The changes which occur in electronic properties as the system length scale is reduced are related mainly to the increasing influence of the wave-like property of the electrons (quantum mechanical effects) and the scarcity of scattering centres. As the size of the

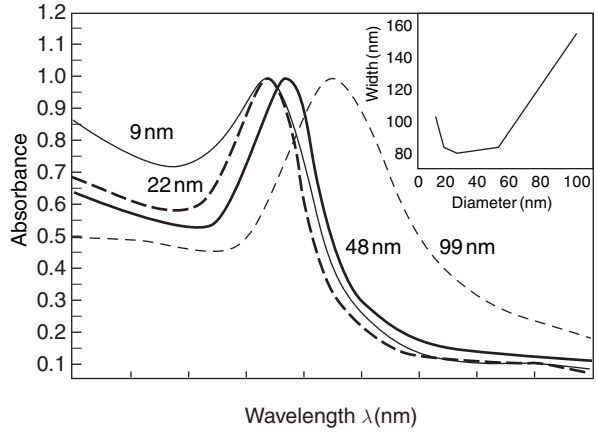


Figure 1.17 Size dependence of the optical absorption wavelength for gold nanoparticles and (inset) the corresponding value of the full width at half maximum (FWHM) of the absorption peak

system becomes comparable with the de Broglie wavelength of the electrons, the discrete nature of the energy states becomes apparent once again, although a fully discrete energy spectrum is only observed in systems that are confined in all three dimensions. In certain cases, conducting materials become insulators below a critical length scale, as the energy bands cease to overlap. Owing to their intrinsic wave-like nature, electrons can tunnel quantum mechanically between two closely adjacent nanostructures, and if a voltage is applied between two nanostructures which aligns the discrete energy levels in the DOS, *resonant tunnelling* occurs, which abruptly increases the tunnelling current.

In macroscopic systems, electronic transport is determined primarily by scattering with phonons, impurities or other carriers or by scattering at rough interfaces. The path of each electron resembles a random walk, and transport is said to be *diffusive*. When the system dimensions are smaller than the electron mean free path for inelastic scattering, electrons can travel through the system without randomization of the phase of their wavefunctions. This gives rise to additional localization phenomena which are specifically related to phase interference. If the system is sufficiently small so that all scattering centres can be eliminated completely, and if the sample boundaries are smooth so that boundary reflections are purely specular, then electron transport becomes purely *ballistic*, with the sample acting as a waveguide for the electron wavefunction.

Conduction in highly confined structures, such as quantum dots, is very sensitive to the presence of other charge carriers and hence the charge state of the dot. These *Coulomb blockade* effects result in conduction processes involving single electrons and as a result they require only a small amount of energy to operate a switch, transistor or memory element.

All these phenomena can be utilised to produce radically different types of components for electronic, optoelectronic and information processing applications, such as resonant tunnelling transistors and single-electron transistors. Further details of these concepts and their applications are given in Chapter 3.

1.3.3.8 Biological systems

Biological systems contain many examples of nanophase materials and nanoscale systems. Biomineralization of nanocrystallites in a protein matrix is highly important for the formation of bone and teeth, and is also used for chemical storage and transport mechanisms within organs. Biomineralization involves the operation of delicate biological control mechanisms to produce materials with well-defined characteristics such as particle size, crystallographic structure, morphology and architecture. Generally complex biological molecules such as DNA have the ability to undergo highly controlled and hierarchical self-assembly, which makes them ideal for the assembling of nanosized building blocks. Self-assembly is discussed in Section 1.4.3.1 and in detail in Chapters 7 and 8. Methods for altering and controlling these interactions and building nanoscale building blocks and assembling nanoscale architectures are also discussed.

Biological cells have dimensions within the range 1–10μm and contain many examples of extremely complex nano-assemblies, including molecular motors, which

are complexes embedded within membranes that are powered by natural biochemical processes (Figure 1.18).

Generally, naturally occurring biological nanomaterials have been refined by evolutionary processes over a long timescale and are therefore highly optimized. We can often use biological systems as a guide to producing synthetic nanomaterials or nanosystems, a process known as *biomimicry*.

[Image not available in this electronic edition.]

Figure 1.18 A transmission electron microscope image of two myosin molecules, each of which has two heads that can bind an actin filament and split a biochemical compound called ATP to provide chemical energy for motion. A large head shape change while attached to actin causes muscle contraction and a wide variation in cell motility. A 20 nm scale bar is shown in the image which reproduced from The Journal of Cell Biology, permission of the Rockefeller University Press and courtesy of Professor John Trinick, University of Leeds

1.4 FABRICATION METHODS

Nanostructures can be made in numerous ways. As highlighted in Section 1.1.3, a broad classification divides methods into either those which build from the bottom up, atom by atom, or those which construct from the top down using processes that involve the removal or reformation of atoms to create the desired structure. The two approaches are schematically represented in Figure 1.19.

In the bottom-up approach, atoms, molecules and even nanoparticles themselves can be used as the building blocks for the creation of complex nanostructures; the useful size of the building blocks depends on the properties to be engineered. By altering the size of the building blocks, controlling their surface and internal chemistry, and then controlling their organization and assembly, it is possible to engineer properties and functionalities of the overall nanostructured solid or system. These processes are essentially highly controlled, complex chemical syntheses. On the other hand, top-down approaches are inherently simpler and rely either on the removal or division of bulk material, or on the miniaturization of bulk fabrication processes to produce the desired structure with the appropriate properties. When controlled, both top-down and bottom-up methods may be viewed as essentially different forms of microstructural engineering. As shown in Figure 1.19, in terms of scale, biological processes are essentially intermediate between top-down and bottom-up processes, however in reality they usually constitute complex bottom-up processes. A brief overview of some of the more common fabrication methods for nanostructures is given below; further details are discussed in subsequent chapters.

1.4.1 Top-down processes

Top-down processes are effectively examples of solid-state processing of materials.

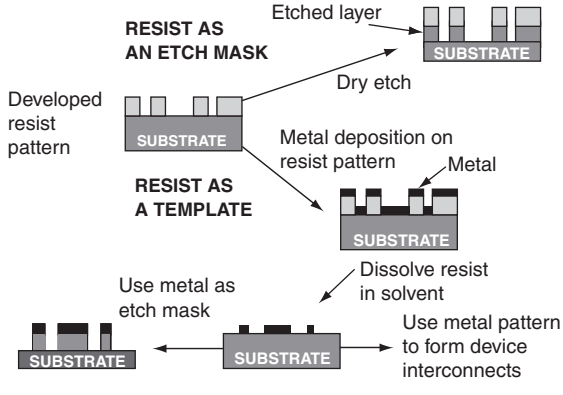


Figure 1.19 Schematic representation of the top-down and bottom-up processes and their relationship to biological processes and structures

1.4.1.1 Milling

One nanofabrication process of major industrial importance is high-energy ball milling, also known as mechanical attrition or mechanical alloying. As shown in Figure 1.20, coarse-grained materials (usually metals but also more recently ceramics and polymers) in the form of powders are crushed mechanically in rotating drums by hard steel or tungsten carbide balls, usually under controlled atmospheric conditions to prevent unwanted reactions such as oxidation. This repeated deformation can cause large reductions in grain size via the formation and organization of grain boundaries within the powder particles. Different components can be mechanically alloyed together by cold welding to produce nanostructured alloys. A nanometre dispersion of one phase in another can also be achieved. Microstructures and phases produced in this way can often be thermodynamically metastable. The technique can be operated at a large scale, hence the industrial interest.

Generally any form of mechanical deformation under shear conditions and high strain rates can lead to the formation of nanostructures, since energy is being continuously pumped into crystalline structures to create lattice defects. The severe plastic deformation that occurs during machining, cold rolling, drawing, cyclic deformation or sliding wear has also been reported to form nanostructured material.

1.4.1.2 Lithographic processes

Conventional lithographic processes are akin to the emulsion-based photographic process and can be used to create nanostructures by the formation of a pattern on a substrate via the creation of a resist on the substrate surface. One lithographic method uses either visible or ultraviolet (UV) light, X-rays, electrons or ions to project an image containing the desired pattern onto a surface coated with a photoresist material; this method requires the prior fabrication of an absorbing mask through which the parallel radiation passes before shadowing onto the photoresist. Alternatively, primary patterning (or direct writing) of the resist is possible using a focused electron, ion or possibly X-ray beam; here either the focused beam or the resist itself is scanned according to the desired pattern design. These different techniques are generally termed photolithography, X-ray lithography, electron beam lithography or ion beam lithography depending on the radiation employed.

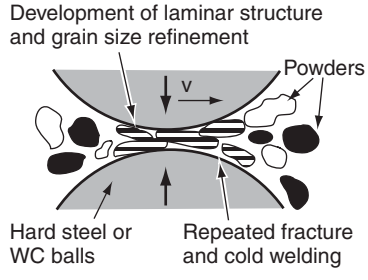


Figure 1.20 Schematic representation of the mechanical alloying process

The resist material, typically a polymer, metal halide or metal oxide, is chemically changed during irradiation, often altering the solubility or composition of the exposed resist. In a positive resist, irradiated areas are dissolved as they are more soluble than unexposed material, whilst in a negative resist irradiated areas are insoluble. As shown in Figure 1.21, the resist can then be used as a template, or as an etch-mask, for subsequent deposition onto, or etching of, the underlying substrate. Resists based on self-assembled monolayers on substrate surfaces are currently being investigated.

The pattern transfer processes which utilize the patterned resist may be divided into solution-based wet chemical etching procedures, dry etching in a reactive plasma, doping using ion implantation techniques, or thin film deposition. Dry etching is the collective term for a range of techniques such as reactive ion etching (RIE) and chemically assisted ion beam etching (CAIBE), which are used extensively for high-resolution pattern transfer; both methods produce, either directly or indirectly, reactive ion species which combine with the elements in the substrate material to form volatile reaction products which are removed into a vacuum system.

Lithographic techniques are very heavily used in the semiconductor processing industry for the fabrication of integrated circuits, optoelectronic components, displays and electronic packaging. The important considerations here are not only the uniformity and reproducibility of the fabrication process, but also the time required to pattern a given area of a planar device, which is summarized in terms of the areal throughput (in microns per hour). A further consideration for nanostructures, particularly for the production of integrated circuits and microelectromechanical systems (MEMS), is the ultimate resolution of the lithographic technique. Fundamentally, the wavelength of the radiation used in the lithographic process determines the detail in the resist and hence the final planar nanostructure; additional considerations may involve the limitations of the projection optics and the nature of the interaction of the radiation with the resist material. Typically the resolution ranges from a few hundred nanometres for optical techniques to tens of nanometres for electron beam techniques. Phenomenologically, throughput and resolution of lithographic techniques broadly follow a power-law dependence; the resolution is approximately equal to $23A^{0.2}$, where A is the areal throughput. Optical, UV and X-ray lithography are fast, parallel exposure techniques capable of micron and submicron resolution, whereas electron and ion beam methods provide nanometre resolution, but

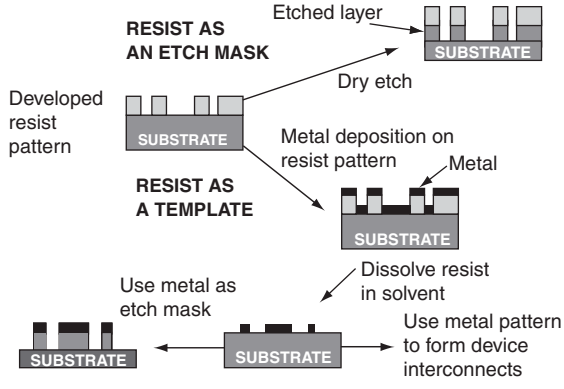


Figure 1.21 Schematic representation of various types of photolithographic process

are, at present, considerably slower as they involve serial exposure using scanned focused probes controlled by computer-aided design (CAD) software. Parallel projection electron beam systems (such as the Scalpel method[4]) and X-ray proximity printing methods are currently being developed for high-resolution, large-throughput mask fabrication. A further area which may be developed in the future is the use of massive arrays of individually controllable atomic force microscope (AFM) or scanning tunnelling microscope (STM) tips for parallel atom manipulation on planar substrates; these are essentially MEMS devices integrated with microprocessor control for the direct patterning of resists (e.g., self-assembled monolayers) or even substrates at reasonable throughputs.

Soft lithography techniques (Figure 1.22) pattern a resist by physically deforming (or embossing) the resist shape with a mould or stamp, rather than by modifying the resist chemical structures with radiation as in conventional lithography. Additionally the stamp may be coated with a chemical that reacts with the resist solely at the edges of the stamp. These methods circumvent many of the resolution limitations inherent in conventional lithographic processes that arise due to the diffraction limit of the radiation, the projection or scanning optics, the scattering process and the chemistry within the resist material. Ultimately, nanoimprinting should represent a cheaper process for mass production. Currently, these soft lithography techniques can produce patterned structures in the range 10 nm and above. One of the main limitations on resolution arises from plastic flow of the polymeric materials involved. Master moulds may be fabricated using either conventional lithographic techniques, micromachining or naturally occurring surface relief on the substrate materials. Once the master mould is formed, a low molecular weight prepolymer is poured and then cured to yield a polydimethylsiloxane (PDMS) elastomeric 'stamp' which may simply be peeled off the master. The relief features that result on the bottom of the PDMS stamp are an inverted replica of those on the master. They may be inked with a molecule of interest (typically by wiping a Q-tip, dampened with the molecule of interest, across their surfaces) and simply placed against a substrate, transferring molecules in a pattern that reflects the features of the stamp. Elastomeric PDMS stamps have been used

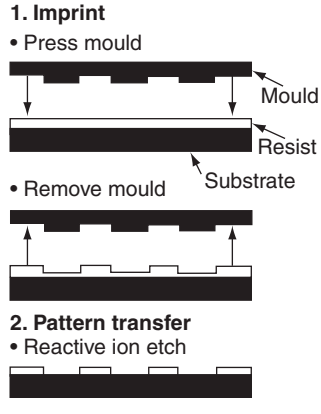


Figure 1.22 Schematic representation of the soft lithographic process which patterns a resist using a stamp followed, if required, by subsequent etching

[4] J. A. Liddle *et al.*, *J Vac. Sci. Technol. B* **19**, 476 (2001).

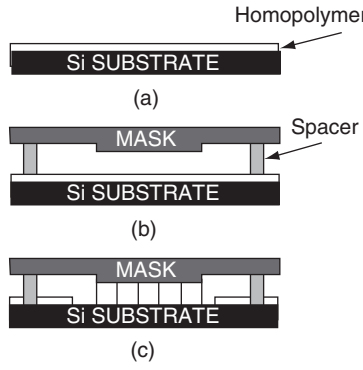


Figure 1.23 Schematic representation of the self-assembly process of a surface polymer film lithographically induced by a mask

to impressive effect as conformal contact masks for phase-shifting photolithography, facilitating the fabrication of structures as small as 50 nm in photoresist. Channels formed in PDMS stamps have also been used to mould three-dimensional structures, a process known as micromoulding in capillaries (MIMIC).

Another variation of this technique is lithographically induced self-assembly (Figure 1.23), in which a mask is used to induce and control, via electrostatic and hydrodynamic instabilities, the formation of periodic supramolecular pillar arrays on a thin polymer melt initially deposited flat onto a substrate. Alternatively, microcontact printing of self-assembled monolayers (SAMs) can be achieved by transfer of a SAM 'ink' onto a substrate surface from a patterned 'stamp' possessing elastomeric properties so as to achieve intimate contact with the surface. A related, and rapidly emerging technique is direct inkjet printing of (currently) micron-scale structures onto surfaces.

1.4.1.3 Machining

Lithographic techniques, whilst being a parallel batch processes, essentially consist of a two-dimensional chemical or mechanical patterning of the surface of a material. More intricate three-dimensional patterning of a material can be achieved by techniques analogous to more conventional machining. Currently the resolution limits of conventional machining are of the order of 5 μm, however, in recent years focused ion beams (FIBs) and high-intensity lasers have been used to directly pattern or shape materials at micron and submicron levels. Figure 1.24 shows a submicron structure produced in a FIB. The length scale achievable in such direct sculpture of materials is not only determined by the spot size and power (or current) density of the radiation employed but also by the nature of the material itself and the removal process.

Modern FIB machines have very accurate sample manipulation stages and are not only capable of material removal, via sputtering of surface substrate atoms by the computer-controlled scanning of focused ion beams produced by liquid metal gallium ion guns, but also of material deposition using the interaction of the ion beam with reactive gases from localized microinjectors. Modern dual-beam FIBs can

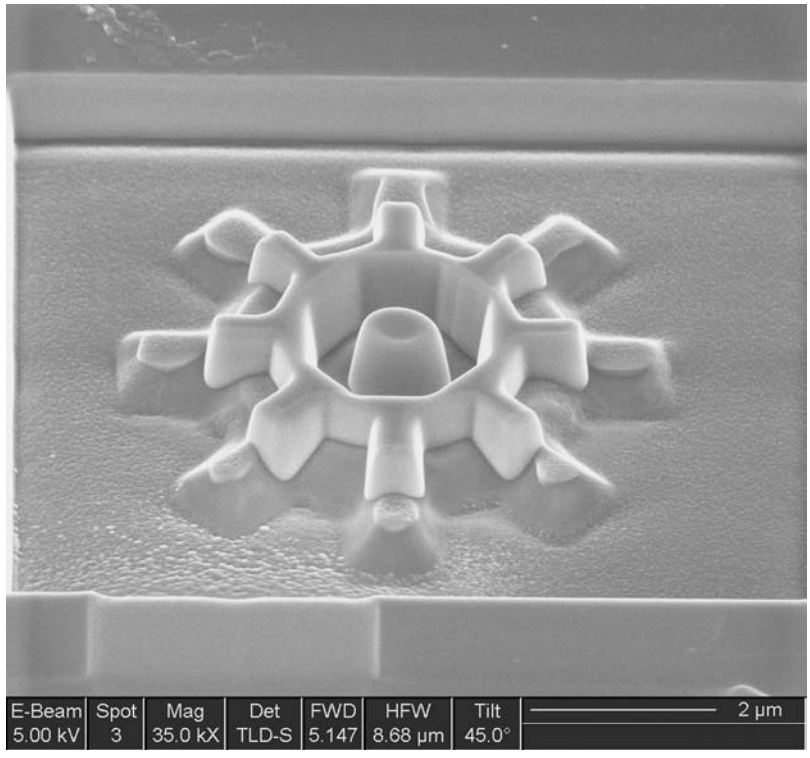


Figure 1.24 Scanning electron microscope image of a multilevel gear structure created by focused ion beam sputtering of silicon. Image courtesy of FEI

simultaneously view the etching or deposition procedures using a separate scanning electron microscope column. As well as inorganic materials, there is increasing interest in FIB patterning of biological structures. Thin sample preparation of specimens for transmission electron microscopy (Chapter 2) is routinely undertaken in a FIB.

1.4.2 Bottom-up processes

Bottom-up processes effectively encompass chemical synthesis and/or the highly controlled deposition and growth of materials. Chemical synthesis may be carried out in either the solid, liquid or gaseous state. Solid-state synthesis usually involves an iterative procedure of bringing solid reaction precursors into intimate contact by mixing and grinding and then promoting atomic diffusion processes via heat treatment at high temperatures to form a reaction product. Such elevated temperatures often lead to rapid grain growth and ultimately a final product with a relatively large grain size unless grain growth inhibitors are present. Consequently, true nanoscale systems are difficult to obtain via solid-state routes, apart from perhaps the controlled growth of second phase nanoscale precipitates within a primary matrix such as in precipitation-strengthened steels, aluminium alloys or other nanocomposite systems. These methods are discussed

in Chapter 5. Here we focus primarily on liquid and gas phase fabrication routes. Diffusion of matter in the liquid and gas phases is typically many orders of magnitude greater than in the solid phase; therefore these synthetic methods can be implemented at much lower temperatures, thus inhibiting unwanted grain growth and so resulting in true nanoscale systems.

1.4.2.1 Vapour phase deposition methods

Vapour phase deposition can be used to fabricate thin films, multilayers, nanotubes, nanofilaments or nanometre-sized particles. The general techniques can be classified broadly as either physical vapour deposition (PVD) or chemical vapour deposition (CVD).

PVD involves the conversion of solid material into a gaseous phase by physical processes; this material is then cooled and redeposited on a substrate with perhaps some modification, such as reaction with a gas. Examples of PVD conversion processes include thermal evaporation (such as resistive or electron beam heating or even flame synthesis), laser ablation or pulsed laser deposition (where a short nanosecond pulse from a laser is focused onto the surface of a bulk target), spark erosion and sputtering (the removal of a target material by bombardment with atoms or ions).

One example of a PVD technique is vapour phase expansion, which relies on the expansion of a high-pressure vapour phase through a jet into a low-pressure ambient background to produce supersaturation of the vapour. Flow rates can approach supersonic speeds and the process can lead to the nucleation of extremely small clusters, ranging from a few atoms upwards, which can be analysed using mass spectrometry. Although the quantities of such clusters are low, these can be of great use for studying the physics of small, usually metallic, clusters, which often are composed of magic numbers of atoms due to their very stable geometric and electronic configurations. A schematic diagram of such a system is shown in Figure 1.25.

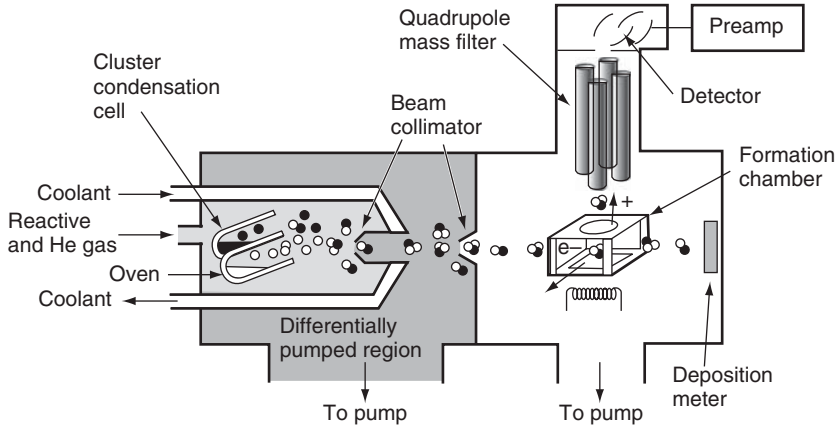


Figure 1.25 Schematic diagram of cluster formation via vapour phase expansion from an oven source

Another PVD process is direct gas phase condensation. Here a material, often a metal, is evaporated from a temperature-controlled crucible into a low-pressure, inert gas environment – ultra-high vacuum (UHV) inert gas condensation. The metal vapour cools through collisions with the inert gas species, becomes supersaturated and then nucleates homogeneously; the particle size is usually in the range 1–100 nm and can be controlled by varying the inert gas pressure. Particles can be collected on a cold finger cooled by liquid nitrogen, scraped off and compacted to produce a dense nanomaterial. Figure 1.26 shows the experimental apparatus. Further details are given in Section 5.3.3.

CVD involves the reaction or thermal decomposition of gas phase species at elevated temperatures (typically 500–1000 °C) and subsequent deposition onto a substrate. A simple example is aerosol spray pyrolysis involving aqueous metal salts which are sprayed as a fine mist, dried and then passed into a hot flow tube where pyrolysis converts the salts to the final products. Since materials are mixed in solution, homogeneous mixing at the atomic level is possible and pyrolysis at relatively low temperatures provides particles in the size range 5–500 nm.

Several CVD processes employ catalysts to enhance the rates of certain chemical reactions. When the catalyst is in the form of a nanometre-sized dispersion of particles, this can often provide a templating effect; for example, in the prodution of carbon nanotubes using the decompostion of ethyne or ethene with hydrogen, Fe-, Co- or Ni-based

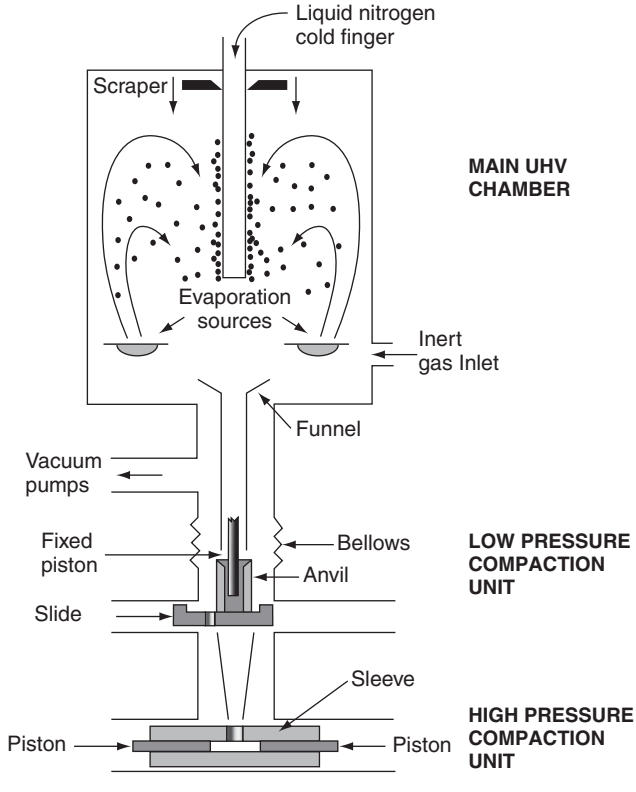


Figure 1.26 Schematic diagram of an inert gas condensation apparatus

catalysts are employed. The size and size distribution of the catalyst particles may determine the internal diameter of the nanotube.

1.4.2.2 Plasma-assisted deposition processes

The use of plasmas (i.e., ionized gases) during vapour deposition allows access to substantially different chemical and physical processes and also higher-purity final materials relative to the conventional PVD and CVD processes described above. There are several different types of plasma deposition reactor for plasma-assisted PVD.

DC glow discharge

DC glow discharge involves the ionization of gas atoms by electrons emitted from a heated filament. The gas ions in the plasma are then accelerated to produce a directed ion beam. If the gas is a reactive precursor gas, this ion beam is used to deposit directly onto a substrate; alternatively an inert gas may be used and the ion beam allowed to strike a target material which sputters neutral atoms onto a neighbouring substrate (Figure 1.27).

Magnetron sputtering

Magnetron sputtering involves the creation of a plasma by the application of a large DC potential between two parallel plates (Figure 1.28). A static magnetic field is applied near a sputtering target and confines the plasma to the vicinity of the target. Ions from the high-density plasma sputter material, predominantly in the form of neutral atoms, from the target onto a substrate. A further benefit of the magnetic field is that it prevents secondary electrons produced by the target from impinging on the substrate and causing heating or damage. The deposition rates produced by magnetrons are high enough ($\sim 1\,\mu m/min$) to be industrially viable; multiple targets can be rotated so as to produce a multilayered coating on the substrate.

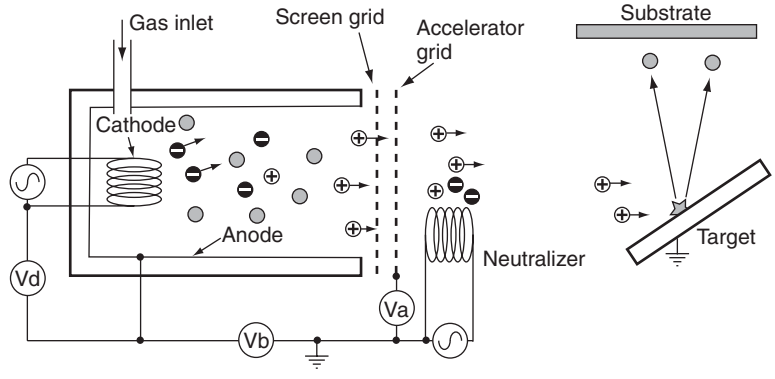


Figure 1.27 Schematic diagram of a DC glow discharge apparatus in which gas atoms are ionized by an electron filament and either deposit on a substrate or cause sputtering of a target

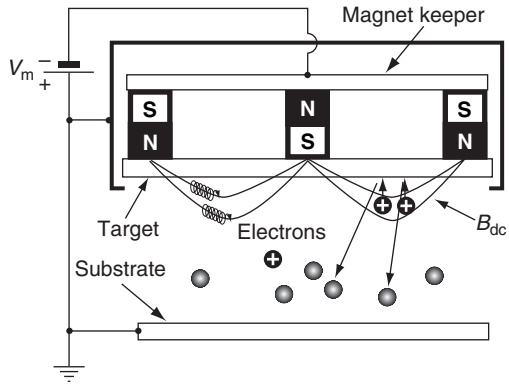


Figure 1.28 Schematic diagram of a magnetron sputtering apparatus that uses a magnetic field to contain ionized gas molecules close to a sputtering target, giving large deposition rates

Vacuum arc deposition

Vacuum arc deposition involves the initiation of an arc by contacting a cathode, constructed from the target material, with an igniter attached to an anode; this generates a low-voltage, high-current arc, which is self-sustaining. The arc ejects predominantly ions and large, micrometre-sized droplets from a small area on the cathode. The ions in the arc are accelerated towards a substrate and deflected using a magnetic field if desired; any large particles being filtered out before deposition (Figure 1.29). A vacuum cathodic arc can operate without a background gas under high-vacuum conditions, however, reactive deposition can be achieved by introducing a background gas such as nitrogen. The high ion energy produces dense films even at low substrate temperatures and consequently arc technology is commonly used for the deposition of hard coatings. As well as DC sources, plasmas can also be produced using radio (MHz) frequency and microwave (GHz) frequency power; these methods have the advantage of being able to provide higher current densities and higher deposition rates.

In the plasma-assisted PVD processes described above, all vapour phase species originate from a solid target. Instead, plasma-enhanced CVD (PECVD) employs gas phase precursors that are dissociated to form molecular fragments which condense to form thin

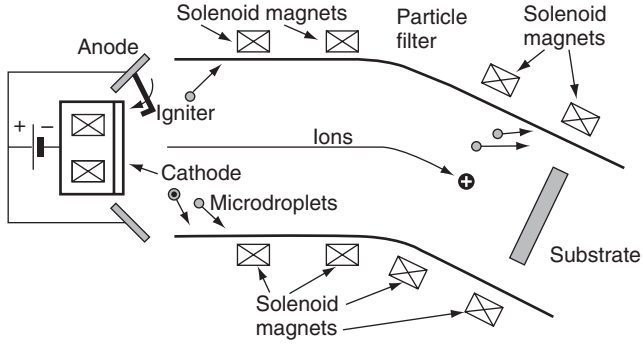


Figure 1.29 Schematic diagram of a vacuum arc deposition apparatus

films or nanoparticles. The dissociation temperatures required for PECVD tend to be much lower than for conventional CVD processes due to the high energy of the plasma, and this may be of importance for deposition on sensitive substrates such as semiconductors and polymers.

1.4.2.3 MBE and MOVPE

A molecular beam epitaxy (MBE) machine is essentially an ultra-high-precision, ultra-clean evaporator, combined with a set of in-situ tools, such as RHEED and Auger spectroscopy (see Chapter 2), for characterization of the deposited layers during growth. The reactor consists of an ultra-high-vacuum chamber (typically better than $\sim 5 \times 10^{-14}$ atmospheric pressure) of approximately 1.5 m diameter (Figure 1.30).

MBE is a growth technique in which epitaxial, single atomic layers ($\sim 0.2-0.3$ nm) are grown on a heated substrate under UHV conditions, using either atomic or molecular beams evaporated from effusion sources with openings directed towards a heated substrate usually consisting of a thin (~ 0.5 mm) wafer cut from a bulk single crystal. The sources can be either solid or gaseous and an MBE machine will typically have an array of multiple sources, which can be shuttered to allow layered, alternating heterostructures to be produced. Semiconductor quantum wells, superlattices and quantum wires (Chapter 3) and metallic or magnetic multilayers for spin valve structures (Chapter 4) are deposited using this technique.

Standard MBE uses elements in a very pure form as solid sources contained within a number of Knudsen cells. In operation the cells are heated to the temperature at which the elements evaporate, producing beams of atoms which leave the cells. The beams intersect at the substrate and deposit the appropriate semiconductor, atomic layer by atomic layer. The substrate is rotated to ensure even growth over its surface. By operating mechanical

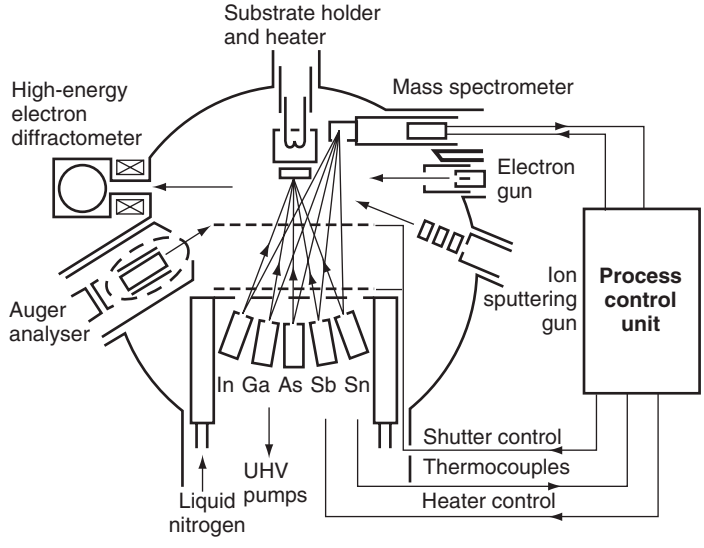


Figure 1.30 Schematic diagram of a molecular beam epitaxy thin film deposition system

shutters in front of the cells, it is possible to control which semiconductor or metal is deposited. For example, opening the Ga and As cell shutters results in the growth of GaAs. Shutting the Ga cell and opening the Al cell switches the growth to AlAs. As the shutters can be switched rapidly, in comparison to the rate at which material is deposited, it is possible to grow very thin layers exhibiting very sharp interfaces. Other effusion cells contain elements required for doping, and it is possible to monitor the growth by observing the electron diffraction pattern produced by the surface (Chapter 2). MBE can also be performed using gaseous sources, and this is often termed chemical beam epitaxy (CBE). When the sources are metallorganic compounds, the process is known as metallorganic MBE (MOMBE).

The second main epitaxial technique is metallorganic vapour phase epitaxy (MOVPE), also known as metallorganic CVD (MOCVD). In this technique the required elements are transported as components of gaseous compounds such as metal alkyls and non-metal hydrides (e.g., trimethylgallium $(CH_3)_3Ga$, arsine AsH_3) to a suitable chamber where they flow over the surface of a heated substrate. These compounds break down and react so as to deposit the relevant semiconductor, with the remaining waste gases being removed from the chamber. MOVPE is used extensively for the production of compound semiconductor (e.g., III–V) thin films, such as the production of GaAs and AlGaAs layers using trimethyl gallium (TMG):

$$Ga(CH_3)_3 + AsH_3(g) \rightarrow GaAs(s) + 3CH_4(g), \tag{1.24a}$$

$$\begin{aligned} x\mathrm{Al}(\mathrm{CH_3})_3(\mathrm{g}) + (1-x)\mathrm{Ga}(\mathrm{CH_3})_3(\mathrm{g}) + \mathrm{AsH_3(g)} \\ \rightarrow \mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As(s)} + 3\mathrm{CH_4(g)}. \end{aligned} \tag{1.24b}$$

Valves in the gas lines allow gas switching, thereby enabling layered structures to be deposited. However, since rapid switching of a gas is more difficult than for an atomic beam, and because of residual gas adsorption on the walls of the feed pipework, the interfaces between adjacent layers grown by the gas-source epitaxy methods are not as abrupt as those attainable with solid-source MBE. For the same reason, solid-source MBE is capable of growing thinner layers, and the slow growth rates involved (of order 1 μm per hour) generally give rise to higher precision in layer thicknesses compared to MOVPE and MOMBE. On the other hand, the high growth rate of MOVPE is better suited for commercial production, especially for semiconductor lasers where thick cladding or waveguiding layers must be grown in a reasonable timescale, in addition to the thin quantum well layers. MOVPE is also easier to scale up to larger systems, permitting simultaneous growth on multiple substrates. Gas source MBE, including MOMBE, generally operates with speeds and precisions which are intermediate between those of solid-source MBE and MOVPE. All the metallorganic epitaxy techniques have significant safety implications as the gases used are highly toxic.

By definition, the aim of molecular beam epitaxy is to produce a perfect single-crystal structure whose morphology and lattice spacing exactly match those of the substrate material. This requires the equilibrium lattice constant of all the different epitaxial layers to be closely matched to that of the substrate, otherwise dislocations will form. (The epitaxial growth of strained semiconductors is discussed in Chapter 3.) For some

combinations of layers, a suitable substrate cannot be identified, in which case an initial buffer layer is deposited to serve as a 'virtual' substrate. In some material systems the deposited layers are sometimes annealed to produce an atomically smooth surface prior to further deposition of another material.

MBE growth is generally far from equilibrium and various growth modes can be identified as shown in Figure 1.31. The prevalence of each mode depends on the exact substrate temperature and deposition rate.

- 1D step propagation growth occurs at high temperatures and/or low growth rates. It occurs on vicinal, stepped surfaces when the surface diffusion of the growing species is pronounced. Deposited material diffuses to a step and is incorporated into the laterally growing terrace. This growth mode produces very flat and smooth interfaces.
- 2D nucleation and growth occurs at intermediate temperatures and/or growth rates and involves the initial nucleation of islands, which grow and coalesce to form a network of interconnected islands separated by holes. These holes are finally filled and a flat surface is formed. The process then repeats in a cyclical fashion.
- 3D rough growth generally occurs at low temperatures and high deposition rates. Islands nucleate on existing islands and extensive faceting occurs, producing very rough interfaces.

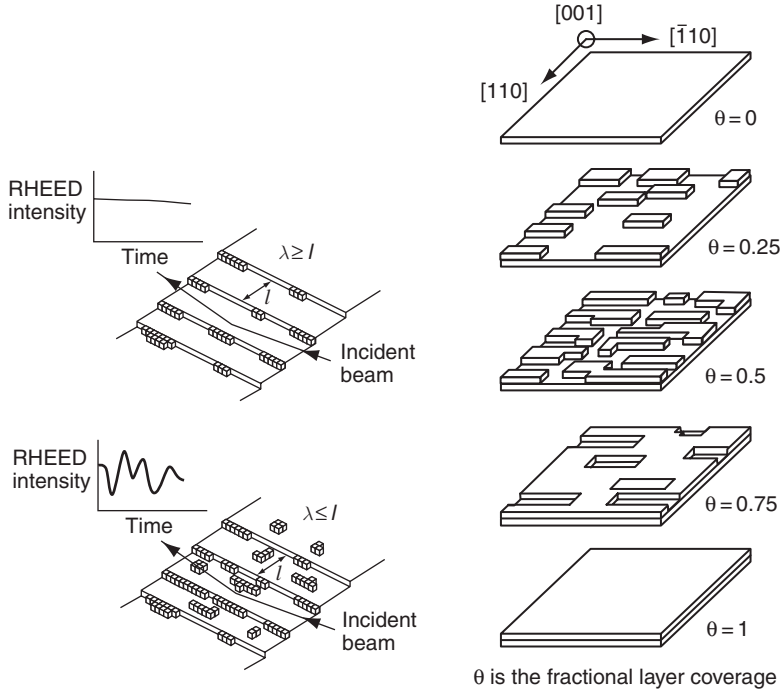


Figure 1.31 Schematic diagram of MBE growth modes: (left) 1D step propagation and (right) 2D nucleation and growth

Because the dopants in both MBE and MOVPE systems exist as additional switchable sources, they may be added in known amounts and distributed over known depth concentration profiles such an abrupt delta-doped layers or extended and/or graded distributions. These processes are discussed in more detail in Chapter 3.

Both MBE and MOVPE are very successful in growing quantum well structures with a sensitive control of layer thickneses. However, it is much more difficult to achieve fine control of lateral dimensions in devices for the production of either quantum wires or quantum dots. One technique for achieving lateral control is by the use of patterning. Patterning can be carried out before or after epitaxial growth (or both) using lithography or etching. In addition, under certain growth conditions, quantum wires or dots may form spontaneously during the epitaxial growth of strained materials. This *self-assembly* growth technique is described in Chapter 3.

1.4.2.4 Liquid phase methods

A variant of many of the PVD processes described above are thermal spraying techniques in which a spray of molten or semi-molten solid particles generated by either an electrical thermal source (e.g., plasma spraying) or by chemical combustion (e.g., flame spraying or high-velocity oxygen fuel spraying) are deposited onto a substrate and undergo rapid solidification. This is extensively used to produce nanocrystalline coatings from nanocrystalline powder, wire or rod feedstocks previously fabricated by the mechanical milling or precipitation routes discussed above.

More generally, liquid phase chemical synthesis involves the reaction of a solution of precursor chemicals in either an aqueous or non-aqueous solvent; these precursors react and/or naturally self-assemble to form a solution supersaturated with the product. Thermodynamically this is an unstable situation and ultimately results in nucleation of the product either homogeneously (in solution) or heterogeneously (on external species such as vessel walls or impurities). Initial nuclei then grow into nanometre-sized particles or architectures according to both thermodynamic and kinetic factors. The nature, size and morphological shape of the precipitated structures can often be controlled by parameters such as temperature, pH, reactant concentration and time. A multicomponent product may require careful control of co-precipitation conditions in order to achieve a chemically homogeneous final product.

1.4.2.5 Colloidal methods

Colloidal methods rely on the precipitation of nanometre-sized particles within a continuous fluid solvent matrix to form a colloidal sol. Generally a finely dispersed system is in a high free energy state, as work has essentially been done to break up the solid, which this is equivalent to the free energy required to produce the increased surface area. The colloidal material will therefore tend to aggregate due to attractive van der Waals forces and lower its energy unless a substantial energy barrier to this process exists. The presence and magnitude of an energy barrier to agglomeration will depend on the balance of attractive and repulsive forces between the particles, as

shown in Figure 1.32. The energy of a colloidal solution which is available for aggregation arises from Brownian motion and is typically of order k_BT. Agglomeration (the formation of strong compact aggregates of nanoparticles) and also flocculation (the formation of a loose network of particles) can be prevented by increasing the repulsive energy term, which is normally short range. This can be achieved by electrostatic or steric stablization, both of which lead to a repulsive contribution to the potential energy.

Surface charges on colloid particles can be easily produced by ionization of basic or acidic groups as the pH is varied. For example, $TiCl_4$ hydrolyses in the presence of a base and forms a colloidal solution of TiO_2; here surface OH^- groups on TiO_2 clusters act as an electrostatic colloid stabilizer. Stabilizing steric effects can be produced by the attachment of a capping layer to the particle surfaces. Additional chemicals are added to the colloidal solution which bind to the cluster surface and block vacant coordination sites, thus preventing further growth. These additives can be polymeric surfactants or stabilizers that attach electrostatically to the surface, or anionic capping agents which covalently bind to the cluster. A further possibility is to precipitate a material within a volume of space defined by a micelle or membrane which acts a barrier to further growth.

Subsequent processing of colloids can involve a number of approaches, including additional colloidal precipitation on particle surfaces to produce a core–shell nanoparticle structure, deposition on substrates to produce quantum dots, self-assembly on substrates as ordered 2D and even 3D arrays, and finally embedding in other media to form a nanocomposite. Colloidal methods are relatively simple and inexpensive and have been extensively used for the production of metal and semiconductor nanocrystals using optimized reactions and reaction conditions. One problem inherent in many colloidal methods is that colloid solutions can often age; that is, the particles can increase their size as a function of time.

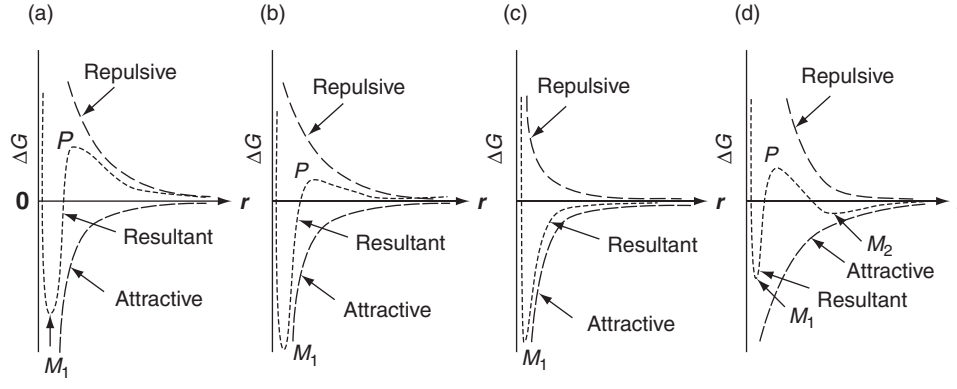


Figure 1.32 Schematic diagram of the possible energy balances between attractive and repulsive interparticulate forces as a function of interparticle separation. In (a) there exists a large energy barrier (P) to strong aggregation of particles (minimum M_1); this is smaller in (b) and absent in (c) owing to reduced long-range repulsive forces. Long-range attractive forces in (d) can lead to weak flocculation of particles (minimum M_2)

1.4.2.6 Sol–gel methods

Sol–gel methods involve a set of chemical reactions which *irreversibly* convert a homogeneous solution of molecular reactant precursors (a sol) into an infinite molecular weight three-dimensional polymer (a gel) forming an elastic solid filling the same volume as the solution. Typically this involves a hydrolysis reaction followed by condensation polymerization, for example:

$$\begin{aligned}&\text{tetraethylorthosilicate (TEOS)}(\mathrm{Si(OC_2H_5)_4})\\&+\text{ethanol}(\mathrm{C_2H_5OH}) \xrightarrow{\text{react and dry}} \mathrm{SiO_2} + \text{other products}\end{aligned} \tag{1.25}$$

Mixtures of precursors can also be used to produce binary or ternary systems, each molecular precursor having its own reaction rate that is dependent on the conditions (e.g., pH, concentration, solvent and temperature). The polymer gel so formed is a 3D skeleton surrounding interconnected pores, and this can be dried and shrunk to form a rigid solid form. In a single precursor component system, the final material can be designed to have interconnected nanoscale porosity and hence a high surface area. For example, nanoporous silica can have a variable percentage porosity that depends on the precursor and the solvent employed.

As well as true molecular precursors, precursors containing nanometre-sized particles can be used, such as aqueous colloidal sols. The transformation to a gel is most often achieved by changing the pH or the concentration of solution; this causes aggregation of the colloidal particles in the sol and results in a skeleton composed of interconnected nanometre particles surrounding pores that are generally bigger than in the corresponding material derived from molecular precursor.

The main benefits of sol–gel processing are the high purity and uniform nanostructure achievable at low temperatures. The method is often used to produce metal-oxide nanomaterials. Further processing usually involves forming the gel using a number of techniques (see below) followed by gentle drying to remove the solvent (this often leads to shrinkage so care is needed to prevent cracks forming during this process). Gels can be cast and moulded to form a microporous preform and dried to produce a monolithic bulk material (e.g., a xerogel or an aerogel) that can be used to form filters and membranes. They can also be spin coated or dipped to produce thin (typically 50–500 nm) films on substrates. These films are used for electronic thin film devices, for wear, chemical or oxidation protection, as well as for their optical properties (e.g., anti-reflection). Alternately, fibres can be drawn from the gel; e.g., silica fibres for light transmission.

The interconnected nanoscale porosity in the dried gel can be filled via incorporation of a second material using techniques such as liquid infiltration or chemical reaction. These materials may then be classed as nanocomposites. Diphasic gels use the initial gel host for the precipitation of a second phase by sol–gel routes. In another variant, organic material can be incorporated as a monomer within an inorganic gel host; the monomers can be subsequently polymerized to form a hybrid material. If a dense rather than a nanoporous material is desired, drying is followed by sintering at higher temperatures. The high surface area leads to rapid densification, which can be accompanied by significant grain growth if temperatures are too high.

1.4.2.7 Electrodeposition

Electrodeposition involves inducing chemical reactions in an aqueous electrolyte solution via use of an applied voltage: methods may be classified as either anodic or cathodic processes. In addition to the simple deposition of metallic thin films, typical examples relevant to nanostructured materials include the deposition of metal oxides and chalcogenides. The electrodeposition process obviously requires an electrically conducting substrate.

In an *anodic* process, a metal anode is electrochemically oxidized in the presence of other ions in solution, which then react together and deposit on the anode. For example:

$$\begin{aligned} &Cd \rightarrow Cd^{2+} + 2e^{-}, \\ &Cd^{2+} + Te^{2-} \rightarrow CdTe. \end{aligned} \tag{1.26}$$

Meanwhile in a *cathodic* process, components are deposited onto the cathode from solution precursors. In the case of metals this process is known as electroplating:

$$\begin{aligned} &Cd^{2+} + 2e^{-} \rightarrow Cd, \\ &HTeO_2^{+} + 4e^{-} + 3H^{+} \rightarrow Te + 2H_2O, \\ &Cd^{2+} + Te^{2-} \rightarrow CdTe. \end{aligned} \tag{1.27}$$

If a metal is oxidized and then undergoes hydrolysis or the local pH at the electrode is changed by liberating hydrogen gas electrochemically and producing OH^{-} ions, then metal oxide materials may be deposited.

Electrodeposition is relatively cheap and can be performed at low temperatures which will minimize interdiffusion if, say, a multilayered thin film material is being prepared. The film thickness can be controlled by monitoring the amount of charge delivered, whereas the deposition rate can be followed by the variation of the current with time. The composition and defect chemistry can be controlled by the magnitude of the applied potential, which can be used to deposit non-equilibrium phases. Pulsing or cycling the applied current or potential in a solution containing a mixture of precursors allows the production of a multilayered material. The potential during the pulse will determine the species deposited whilst the thickness of individual layers is determined by the charge passed. Alternatively, the substrate can be transferred periodically from one electrolytic cell to another. The final films can range in thickness from a few nanometres to tens of microns and can be deposited onto large specimen areas of complex shape, making the process highly suitable for industrial use.

Electrodeposition can also be performed within a nanoporous membrane which serves to act as a template for growth; for example, anodized aluminium has cylindrical nanopores of uniform dimensions (see Figure 1.36) and electrodeposition within this membrane can produce nanocylinders. Deposition on planar substrates can also limit nanocrystal growth and produce ordered arrays; if the growth is epitaxial then any strain due to lattice mismatch between the nanocrystal and the substrate can be growth-limiting. Furthermore, it is possible to modify the surface of substrates (e.g., by STM or AFM) to produce arrays of defects which can act as nucleation sites for the electrodeposition of nanocrystals.

Electroless deposition processes electrons by chemical reactions rather than involve the generation of using an external current as in electroplating and anodization. Unlike electrodeposition, substrates are not required to be electrical conductors, hence biological materials may be coated. One method employs deposition from solutions containing reducing agents where the deposited metal acts as a catalyst for the reduction/deposition process.

1.4.3 Methods for templating the growth of nanomaterials

As discussed in relation to lithographic methods (Section 1.4.1.2), resists can be used as templates for subsequent deposition or etching procedures. Here we consider templating methods utilized principally for wet chemical, bottom-up fabrication methods. One simple method of templating is to precipitate or deposit a material on the outside of another crystal and control the growth. If the template is organic then subsequently it may be washed or burnt out, leaving a porous centre if desired, such as the sol–gel derived silica nanotube templated on tartaric acid shown in Figure 1.1(b). Alternatively a material may be precipitated inside a host material that regulates its growth, as described in detail below.

Small semiconductor nanocrystals, such as CdS_xSe_{1-x}, can be embedded in a glass by the addition of cadmium, sulphur and selenium to a silicate glass melt. The glass is cast and annealed below the melting point, causing nanocrystals to form in the dense glass matrix. However, control of precipitate size can be difficult using this method. Alternatively a porous glass, obtained from sol–gel routes, can be infiltrated at low temperatures with a precursor solution which is then allowed to undergo in-situ precipitation; the pore size of the glass controls the nanoparticle size. Such materials can be used as colour filters and employed in optical devices. Precipitation of general nanocrystals within polycrystalline matrices, such as ferrite (α-Fe) or a ceramic, can be complicated by the crystallization of the matrix; variation in crystallization rates and segregation effects at grain boundaries can lead to an inhomogeneous distribution of precipitates within the microstructure and an inhomogeneous size distribution.

Crystalline hosts such as zeolites can be used as reaction vessels for nanoparticle formation. Zeolites are a range of aluminosilicates, based on building blocks such as sodalite cages, which contain well-ordered, well-defined, interconnected cavities in three dimensions within the structure, having dimensions of the order of nanometres (Figure 1.33).

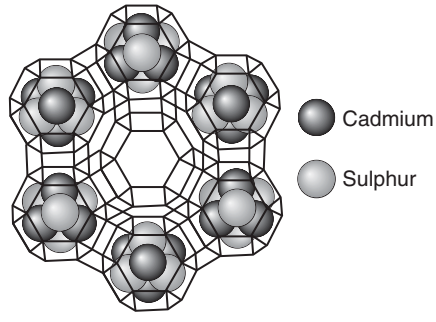


Figure 1.33 The structure of zeolite Y: note isolated Cd_4S_4 units are shown contained within the constituent sodalite cages

The cavities are often charged and are accessed by well-defined windows, which provide a means of transporting reactants to the internal cages. The physical size of the cavities limits the growth of nanoparticles.

Metallic or magnetic nanowires can be formed using a mica or plastic template in which straight, cylindrical tracks have been formed by a high-energy particle beam. Alternatively, metal nanowires can be produced by electrodeposition onto the cleaved edge of an MBE-grown semiconductor heterostructure. The metal preferentially deposits onto the edge of one specific semiconductor layer, yielding a nanowire of a few nanometres in thickness. This approach is an example of the templated growth of a quasi-1D system on a quasi-2D system.

Nanoparticle growth can be regulated within self-organized, biological and synthetic organic membrane assemblies; e.g., micelles, microemulsions, liposomes and vesicles (Figure 1.34). The molecules of these assemblies have a polar head group and a non-polar hydrocarbon tail, which self-assemble into membrane structures in an aqueous environment. Aqueous and reverse micelles have diameters in the range 3–6 nm, whereas microemulsions possess diameters of 5–100 nm. Liposomes and vesicles are closed bilayer aggregates formed from either phospholipids (liposomes) or surfactants (vesicles). Single bilayer vesicles are 30–60 nm in diameter, but multilamellar vesicles can be as large as 100–800 nm. The membrane structures described above can serve as reaction cages to control the nucleation and growth of particles and also to prevent agglomeration.

Infiltration of cations into the structure of a polymer by ion-exchange methods followed by solution or gas chemical treatment to produce in situ precipitation results in polymer composites with a dispersed nanophase. Alternatively a polymer or monomer solution may be simply mixed with a stable colloid and polymerized and dried to produce a composite material. These techniques are used extensively for semiconductor and electroceramic polymer composites, which may then be spin coated onto a substrate to produce a doped polymer thin film.

The macromolecular structure of organic polymers is defined in terms of four variables: the molecular size of the polymer chains, their composition, the sequence of the respective monomer units in the polymer, and their stereochemistry. Increased

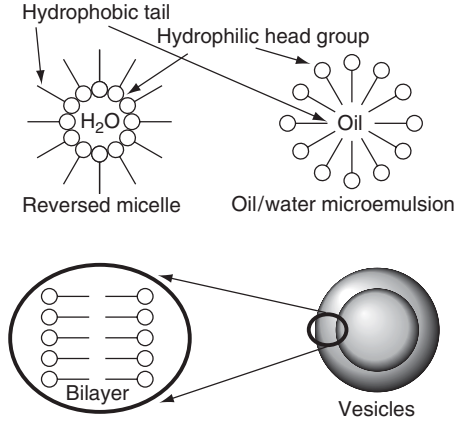


Figure 1.34 Schematic diagram of different self-assembled membrane structures. Image courtesy of IBM Zurich Research Laboratory

control over these variables during polymerization significantly enhances the properties of polymers for industrial application, such as crystallinity and mechanical strength. Usually polymers are synthesized in the laboratory using the iterative (stepwise) addition of selectively activated monomer units to a growing polymer chain, however this becomes increasingly difficult with increasing length and complexity of the polymer molecule. Template-directed polymerization overcomes these problems by using a master template to direct polymer growth in a parallel process from the component monomers; this also provides a verification and proof-reading mechanism for the final fabricated polymer structure. Prime examples of template-directed growth exist in biology; for example, DNA serves as a template for the synthesis of polypeptides through a messenger RNA sequence. Well-defined polypeptides, unlike most synthetic polymers, undergo self-assembly to form ordered three-dimensional structures in solution and in the solid state, as described in the following section.

1.4.4 Ordering of nanosystems

As outlined in Section 1.3, individual nanocrystals or nanoparticles often show very different properties from those of the corresponding bulk material. A further consideration, governing the overall properties of complete nanosystems, is the interaction and coupling between the individual nanocrystalline building blocks. The overall electric, optical, magnetic and transport properties of the nanostructure will then depend not only on the individual nanocrystalline units but also on the coupling and interaction between the nanocrystals, which may be arranged with long-range, translational and even orientational order. A prime example of this phenomenon is a photonic crystal – a periodic array of dielectric particles having separations of the order of the wavelength of light.

There are several ways of producing ordered nanoarchitectures: multicomponent fabrication of multilayered systems by MBE growth of metal or semiconductor heterostructures (Section 1.4.2.3), by processing of particles with core–shell structures, or by fabrication followed by further processing such as etching or focused ion beam milling. Manual or robotic manipulation of nanoparticles on a surface by scanning probe microscopy tips to produce complex structures such as quantum rows and quantum corrals or a molecular abacus (Figure 1.35) with well-defined electronic properties. Initial imaging using AFM or STM is followed by using the tip to push each particle along a desired trajectory, and although overall throughput is slow, parallel arrays of tips in a MEMS device could, in principle, assist larger-scale production. Chemical, physical or geometrical self-assembly or self-organization of atoms, molecules or nanoparticles to form highly ordered nanostructures, as described below.

1.4.4.1 Self-assembly and self-organization

Self-assembly and self-organization of nanostructures is an important area which often bridges the divide between organic and inorganic systems. Many self-assembly processes rely on the self-assembling nature of organic molecules, including complex species such as DNA; these methods are termed chemical or molecular self-assembly. Generally molecular self-assembly is the spontaneous organization of relatively rigid molecules

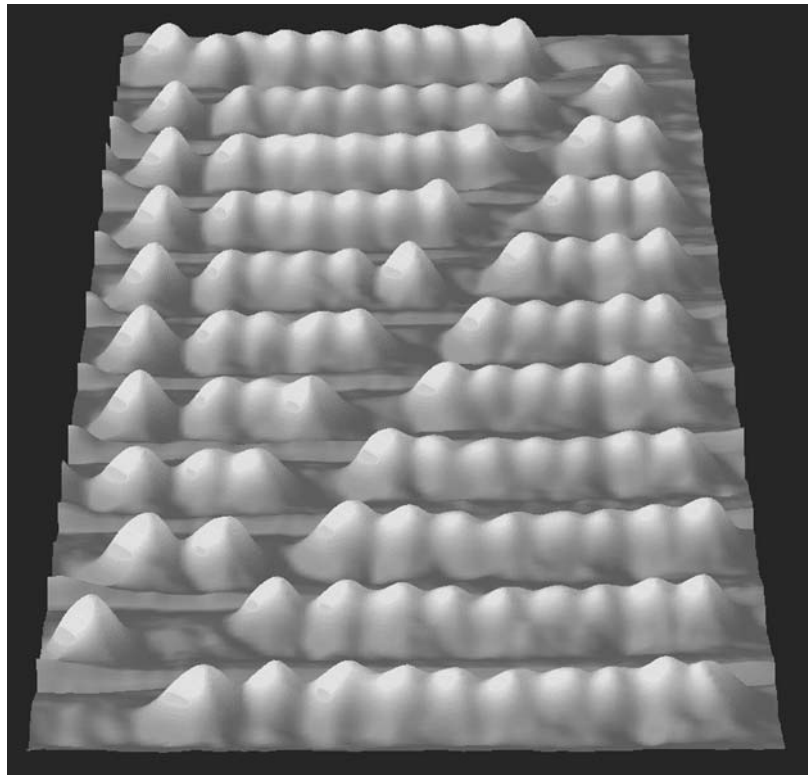


Figure 1.35 A 'molecular abacus'. Image courtesy of IBM Zurich Research Laboratory

into structurally well-defined aggregates via weak, reversible interactions such as hydrogen bonds, ionic bonds and van der Waals bonds. The aggregated structure represents a minimum energy structure or equilbrium phase. Self-assembly is also found throughout biological systems, micelles and liquid crystals and is being increasingly used in synthetic supramolecular chemistry.

Other simpler methods rely on geometric self-organization, in which hard spheres or hard rods will arrange themselves into two- and three-dimensional structures based on packing considerations. For example, solutions of colloidal metal particles can spontaneously order themselves into 2D hexagonally close-packed sheets on substrates. The individual nanoparticles are often encapsulated in a protective organic coating to provide stabilization and a degree of control over the final self-organized structure; for example, gold nanoparticles stabilized with an organic surfactant will self-organize into ordered arrays in which the interparticle spacing depends on the length of the surfactant molecule. Molecular systems, such as rod-like and disc-like liquid crystals, also exhibit geometric self-organization properties.

A variation on geometric self-organization is templated self-organization, in which an ordered nanostructure is formed by deposition of a material around a previously self-organized template. This approach can be used to produce porous metallic structures via electrodeposition on geometrically self-organized polystyrene spheres of submicron dimensions. The spheres are subsequently dissolved to leave a

highly porous, ordered structure. A similar process has been used to produce 3D ordered mesoporous ceramics such as MCM41, which contains an ordered hexagonal pore structure (2–15 nm in size) and has extensive applications in catalysis. Figure 1.36 shows an AFM image of an ordered hexagonal array of pores in an aluminium oxide thin film grown on a pure aluminium substrate by anodization (electrochemically induced oxidation).

More complex self-assembly processes involve the use of self-assembled monolayers (SAMs). SAMs comprise organic molecules whose functionality can be modified by chemical treatment or radiation (e.g., lithography) so that subsequent layers can be selectively attached and used to direct oriented crystal growth. The ends of the molecules are usually terminated with a thiol group to provide good adhesion to a gold substrate. The molecules will order on the surface under given conditions of concentration, pH and temperature. An extensive discussion of molecular self-assembly is given in Chapter 7.

Another important category of self-assembly processes is the self-assembled growth of semiconductor quantum dots. This is achieved via use of a three-dimensional MBE growth mode (the Stranski–Krastonov mode) on a lattice-mismatched substrate (e.g., InAs on GaAs). The strain fields resulting from the lattice mismatch give rise to island growth with a well-defined geometry (e.g., pyramids, cones or lenses) depending on the precise material combination and growth conditions. Unlike colloidal quantum dots, the Stranski–Krastonov dots cannot exist in isolation, and are encapsulated by a subsequently grown semiconductor layer having a sufficient band-gap difference to provide electronic confinement. The quantum dots have no in-plane ordering, but if

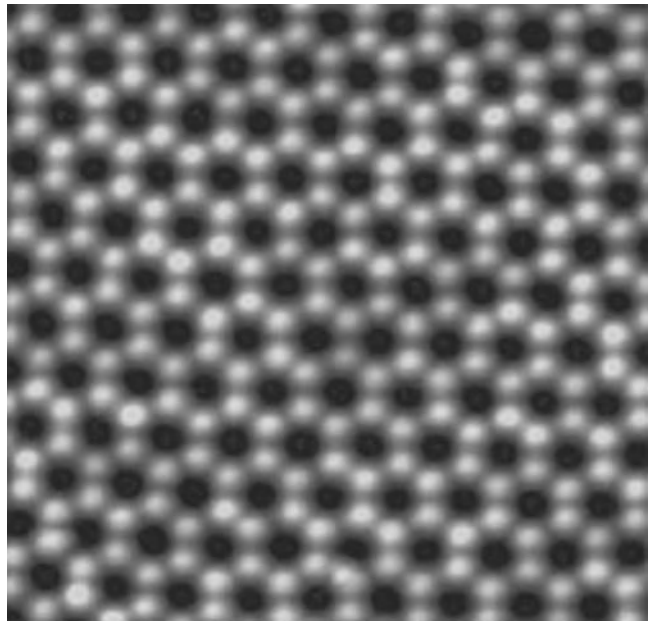


Figure 1.36 A 1 micron square AFM image of an ordered hexagonal array of pores in an anodic alumina film. Image courtesy of Joe Boote and Professor Steve Evans, Department of Physics and Astronomy, University of Leeds

subsequent InAs layers are deposited, the second layer of dots will preferentially nucleate above the first layer. In this way, long columns of quantum dots can be built up with almost perfect vertical ordering. Further details are given in Chapter 3.

1.5 PREPARATION, SAFETY AND STORAGE ISSUES

The preparation of nanoscale systems generally requires a well-regulated laboratory environment since the systems are so small that contamination, even at low levels, becomes a highly important issue. In addition to the use of ultra-high purity reagents, this means that there should also be good control of temperature and humidity, which necessitates the provision of a good air-conditioning system, as well as a well-defined, clean water or solvent supply. Generally the environment should be dust-free and often vibration- and draught-free; it should possess adequate ventilation in the form of fume cupboards, and it should even have facilities for fabrication under a controlled atmosphere environment through use of gloveboxes. Such laboratories are termed clean rooms and will have a standard ISO designation based on the degree of environmental control; this is usually based on the number of submicron particles in a given volume of air. Overall the room is under positive pressure to prevent dust ingress, and incoming air is filtered and recirculated and protective clothing is worn to minimize sources of contamination. The air in a typical room would contain around 100 000 particles per cubic foot, and a fairly clean area may contain around 10 000 particles per cubic foot, formerly known as Class 10 000. Class 10 000 is now more commonly known as ISO Class 7; note that this new standard measures the number of particles per cubic metre and also includes particles smaller than 0.5 microns. ISO Class 7 is used for measurement and packaging areas in integrated circuit fabrication, whereas ISO Class 3 is used during fabrication. Class 1 implies minimal human contact.

Health and safety issues are paramount during any industrial revolution. From the field of environmental particulates there is increasing evidence that nanoscale materials are often highly biologically active, in that they can easily penetrate through biological matter and, owing to their high surface area, will have a large associated chemical reactivity which may result in extensive tissue damage if inhaled or ingested. This means that nanoscale particles and fibres may be biotoxic and, if in doubt, care should be taken when handling or breathing them.

Besides the problem of ageing in nanoparticle solutions, many nanoparticle solutions and more particularly powders can be very reactive (even explosive) due to their high surface area to volume ratio (such as metals in air) and they may need to be stored under an inert atmosphere and at low temperatures. In consolidated nanostructured materials, reactivity and diffusion processes (for example, grain growth) may also be a problem, particularly as the temperature is raised.

BIBLIOGRAPHY

R. R. H. Coombs and D. W. Robinson (Eds), *Nanotechnology in Medicine and Biosciences* (Gordon and Breach, New York, 1996)

A. S. Edelstein and R. C. Cammarata (Eds), *Nanomaterials: Synthesis, Properties and Applications* (Institute of Physics, 1998)

A. N. Goldstein (Ed.), *Handbook of Nanophase Materials* (Marcel Dekker, New York, 1997)
C. Hammond, *The Basics of Crystallography and Diffraction* (Oxford University Press, 1997)
C. C. Koch, *Nanostructured Materials: Processing, Properties and Applications* (Institute of Physics, 2002)
M. Kohler and W. Fritzsche, *Nanotechnology: An Introduction to Nanostructuring Techniques* (Wiley-VCH, Weinheim, 2004)
K. W. Kolasinski, *Surface Science: Foundations of Catalysis and Nanoscience* (Wiley, Chichester, 2002)
C. P. Poole and F. J. Owens, *Introduction to Nanotechnology* (Wiley, New Jersey, 2003)
M. C. Roco, R. S. Williams and P. Alivisatos, *Nanotechnology Research Directions* (Kluwer, Dordecht, 2000)
C. Suryanarayana, J. Singh and F. H. Froes (Eds), *Processing and Properties of Nanocrystalline Materials* (TMS, Warrendale, PA, 1996)
G. Timp (Ed.), *Nanotechnology* (Springer-Verlag, New York, 1999)
Z. L. Wang (Ed.) *Characterisation of Nanophase Materials* (Wiley-VCH, Weinheim, 2000)