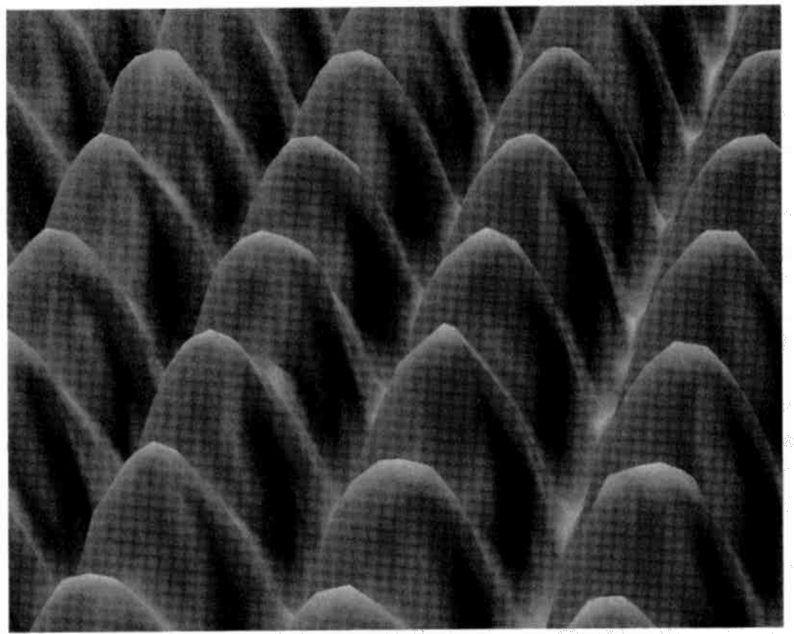
18. For a $1s$ electron in the ground state of hydrogen, determine the expectation value of energy.
19. For an electron in the $(n, l, m) = (2, 0, 0)$ state of hydrogen, determine the expectation value of position.
20. Repeat problem 4.19 for an electron in the $(n, l, m) = (2, 1, 0)$ state of hydrogen.
21. For an electron in the $(n_x, n_y) = (1, 1)$ subband of a metallic quantum wire having $L_x = L_y = 1$ nm, if the total energy is 1 eV, what is the electron's longitudinal (i.e., z-directed) group velocity?
22. A 3 eV electron is to be confined in a square quantum dot of side L. What should L be in order for the electron's energy levels to be well quantized?

Chapter

5

ELECTRONS SUBJECT TO A PERIODIC POTENTIAL — BAND THEORY OF SOLIDS



STM image of a nickel surface. (Image reproduced by permission of IBM Research, Almaden Research Center. Unauthorized use not permitted.)

In the previous chapter, we considered several different environments, including infinite and finite spatial regions, and we solved Schrödinger's equation to determine the possible allowed states for a particle in these environments. In this chapter we continue to use these ideas, but here we consider an electron in a crystalline material, leading to the important *band theory of solids*. It is hard to overestimate the importance of band theory. It turns out that band theory, in conjunction with the Pauli exclusion principle, can be used to explain the fundamental nature of insulators, conductors, and semiconductors. Band theory is also important in understanding semiconductor heterostructures, considered in Chapter 9, which are fundamental structures in nanoelectronics.

Another main point of this chapter is to introduce the concept of effective mass. One can imagine that it would be extremely difficult to account for all of the interactions between an electron and the various particles that make up a material. As described previously, in an exact model this interaction is taken into account via the potential energy term in Schrödinger's equation. However, as we will see, with some simplifying assumptions we can take into account the presence of a crystalline material by simply considering the electron to have a mass that is different from its empty space value. With this idea we can, for example, reconsider the quantum well, quantum wire, and quantum dot examples introduced in the last chapter, this time constructing these structures from a solid material, rather than being simply empty space. The most technically important materials for these applications are semiconductors.

In this chapter, attention is focused on bulk material properties; applications to nanoscale structures made from these materials will be considered in subsequent chapters.

5.1 CRYSTALLINE MATERIALS

The band theory of solids applies only to crystalline materials, strictly speaking, and so it is worthwhile to begin with a brief introduction to the different forms a solid may take.

Solid materials may be classified as *crystalline, polycrystalline, or amorphous.* A crystalline solid has a regular structure, consisting of a periodic array of atoms called the *lattice*. A polycrystalline solid has a well-defined structure in each of many small regions, but each region generally differs from its neighboring regions. In some ways, the opposite of a crystalline material is an amorphous solid, which does not exhibit any sort of regularity. The most common amorphous materials are glass and plastic.

Most materials of interest in electronics have been, traditionally, crystalline materials (e.g., semiconductors, such as silicon and gallium arsenide, and conductors, such as copper and gold). In this chapter, we consider the effect of such a periodic lattice on the behavior of electrons. Although in the past, naturally occurring crystalline materials were used for electronics applications, advances in materials processing technologies have begun to allow synthetic crystalline materials to be developed, engineered specifically to control electronic properties of the material.

Crystal Types. The fundamental property of a crystal is regularity in its atomic structure; the atoms in a crystal are arranged in a regular (periodic) array. To be fairly general, we need to develop two concepts: the idea of a lattice and of a basis.

A lattice is a set of points that form a periodic structure. The simplest lattice, called a *simple cubic* (sc) lattice, consists of points equally spaced at the corners of a three-dimensional cube, as shown in Fig. 5.1, although few materials have this structure.

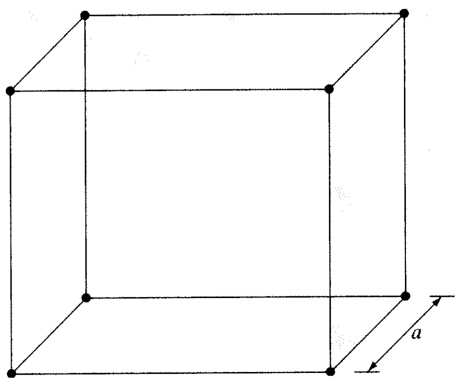


Figure 5.1 Simple cubic lattice. The lattice points are shown by a solid dot, located at the vertex of each corner. All sides have length a.

The lattice is defined by three fundamental vectors, $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$, known as *fundamental translation vectors*, such that the atomic arrangement looks identical when viewed from the point $\mathbf{r}$ and the point

$$\mathbf{r}' = \mathbf{r} + \mathbf{T}, \tag{5.1}$$

where

$$\mathbf{T} = u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + u_3\mathbf{a}_3 \tag{5.2}$$

is called the *crystal translation vector* and where u_1–u_3 are integers. Many translation vectors are possible, and the set of three vectors form a parallelepiped. The parallelepiped with the smallest volume is called the *primitive cell*, constructed from *primitive translation vectors*. The primitive cell has only one lattice point (perhaps shared with other cells), and a crystal can be constructed from repetitions of the primitive cell. However, for envisioning the material to consist of repetitions of cells, the primitive cell is usually not the most convenient to work with, and here we will be primarily interested in other *unit cells*.

Whereas the lattice specifies the periodic arrangement of the crystal, it may not be the case that a single atom is located at each lattice point. A group of atoms called a *basis*, consisting of perhaps many atoms, is such that when the basis is placed at each lattice point the entire crystal is formed. That is, by definition of a lattice, the basis repeats in a periodic manner. The simplest example would be the case when the basis consists of one atom. An example where the basis consists of two different atoms is shown (in two dimensions) in Fig. 5.2.

Just slightly more complex than the sc lattice is the *body-centered cubic* (bcc) lattice, shown in Fig. 5.3. In particular, sodium and tungsten have this structure.

Of more importance for semiconductors is the *face-centered cubic* (fcc) lattice, shown in Fig. 5.4, which is constructed by adding to the simple cubic lattice additional points in the center of each square face. Materials exhibiting this type of lattice are copper (Cu), gold (Au), silver (Ag), nickel (Ni), and, importantly, silicon (Si), gallium arsenide

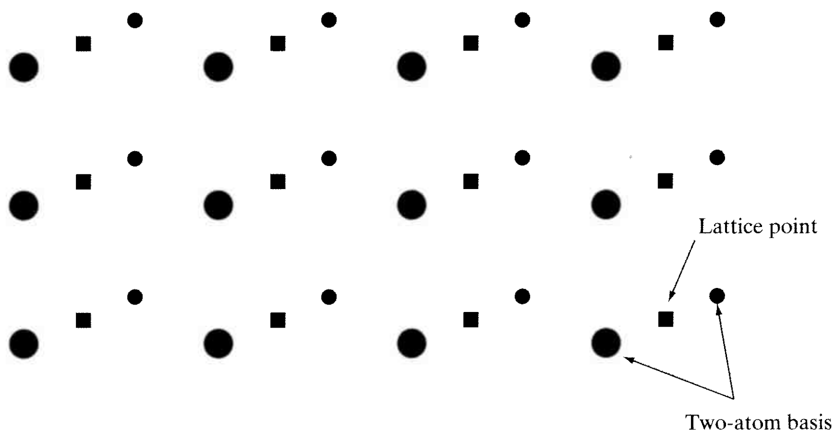


Figure 5.2 Crystal formed by a two-atom basis centered at each lattice point.

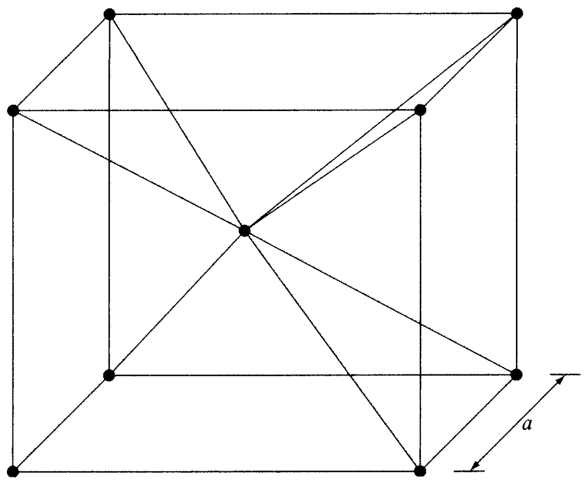


Figure 5.3 Body-centered cubic lattice. The lattice points are shown by a solid dot, located at the vertex of each corner, and in the center of the cube. All sides have length a.

(GaAs), and germanium (Ge). The lattice constant (a) for Si is 0.543 nm, and for Ge, $a = 0.566$ nm. The lattice constants of some other materials are listed in Table V in Appendix B.

To illustrate the idea of a primitive cell and unit cell, Figs. 5.1, 5.3, and 5.4 show what are called the *conventional cells*, which are the structurally simplest unit cells; the lattice constant a is the length of the side of the conventional cell. However, only for the simple cubic lattice is the conventional cell primitive (cell volume is a^3). The primitive cell for the important fcc lattice is shown in Fig. 5.5 (conventional cell volume is a^3, whereas the

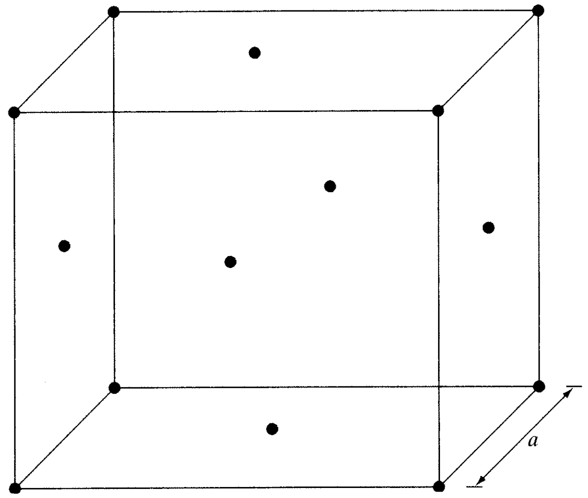


Figure 5.4 Face-centered cubic lattice. The lattice points are shown by a solid dot, located at the vertex of each corner, and in the center of each side of the cube. All sides have length a.

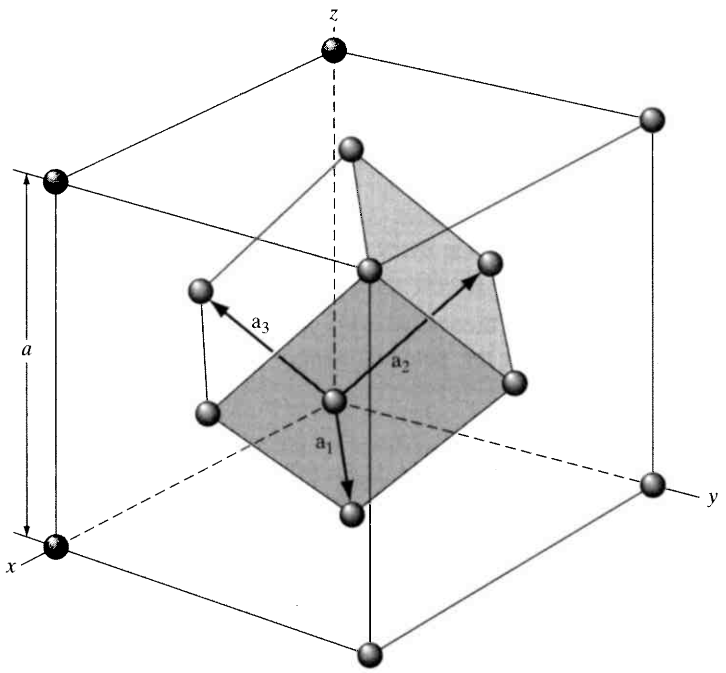


Figure 5.5 Primitive cell for the fcc lattice.

primitive cell volume is $a^3/4$), where the primitive translation vectors are

$$\mathbf{a}_1 = \frac{a}{2}\left(\mathbf{a}_x + \mathbf{a}_y\right), \quad \mathbf{a}_2 = \frac{a}{2}\left(\mathbf{a}_y + \mathbf{a}_z\right), \quad \mathbf{a}_3 = \frac{a}{2}\left(\mathbf{a}_z + \mathbf{a}_x\right). \tag{5.3}$$

Although important semiconductors exhibit the fcc lattice, the basis consists of two atoms. If the two atoms are identical, the material is said to have the *diamond structure*, which occurs for Si, Ge, and carbon (C). If the two atoms in the basis are different, the material is said to have a *zinc blende structure*. Semiconductors such as GaAs and AlAs have this structure, and are often called compound semiconductors (e.g., for GaAs, the two atoms in the basis are gallium and arsenic).

Having established the idea of periodic atomic structure, we consider next how electrons behave in a periodic potential.

5.2 ELECTRONS IN A PERIODIC POTENTIAL

When we want to examine the properties of an electron in a periodic lattice, we need to consider Schrödinger's equation such that the potential energy term $V(\mathbf{r})$ reflects the fact that the electron sees a periodic potential.

Consider a one-dimensional example, where the lattice points are spaced a distance a apart, and we assume an ionized atom (an ion) is located at each lattice point. A liberated electron moves about the material subject to the attractive Coulomb (electrostatic) force between the negatively charged electron and the positively charged ions. The Coulomb force between the electron and a single ion is

$$V(x) = \frac{1}{4\pi\varepsilon_0}\frac{(q_e)(-q_e)}{|x|} \tag{5.4}$$

from basic electrostatic theory, where x is the distance between the electron and the ion, and so the electron has the potential energy depicted in Fig. 5.6. The potential energy is obviously periodic, i.e., $V(x) = V(x + a)$.

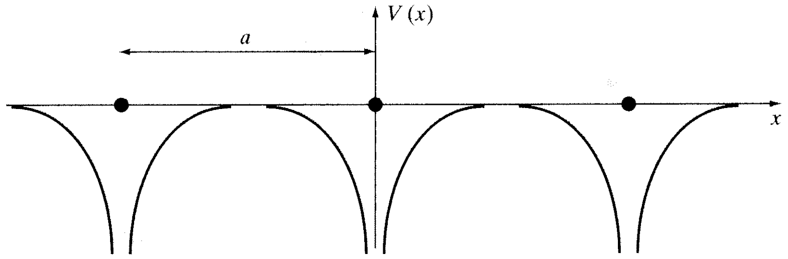


Figure 5.6 Potential $V(x)$ versus position x in a periodic lattice. Solid dots indicate the location of ions, and a is the period of the lattice.

Generalizing to three-dimensions, we find that Schrödinger's equation is

$$\left(-\frac{\hbar^2}{2m_e}\nabla^2 + V(\mathbf{r})\right)\psi(\mathbf{r}) = E\psi(\mathbf{r}), \tag{5.5}$$

where the potential energy term is periodic,

$$V(\mathbf{r}) = V(\mathbf{r} + \mathbf{T}), \tag{5.6}$$

and where $\mathbf{T}$ is the crystal translation vector. There is an important theorem called *Bloch's theorem* that applies to waves in periodic structures in general. (Recall that electrons have wave properties through de Broglie's relation.) For the case of Schrödinger's equation, it states that the solution of (5.5), when the potential is periodic, can be written as the product of a plane wave and a periodic function,

$$\psi(\mathbf{r}) = u(\mathbf{r})\, e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{5.7}$$

where $\mathbf{k}$ is the wavevector to be determined (called the *Bloch wavevector*) and where u is periodic,

$$u(\mathbf{r}) = u(\mathbf{r} + \mathbf{T}). \tag{5.8}$$

Thus,

$$\psi(\mathbf{r} + \mathbf{T}) = u(\mathbf{r} + \mathbf{T})\, e^{i\mathbf{k}\cdot(\mathbf{r}+\mathbf{T})} = u(\mathbf{r})\, e^{i\mathbf{k}\cdot\mathbf{r}} e^{i\mathbf{k}\cdot\mathbf{T}} = \psi(\mathbf{r})\, e^{i\mathbf{k}\cdot\mathbf{T}}. \tag{5.9}$$

It is important to note that the Bloch theorem shows that electrons can propagate through a *perfect* periodic medium without scattering (i.e., without hitting the atoms). One could say, mathematically, that this is because the plane wave part of the solution, $e^{i\mathbf{k}\cdot\mathbf{r}}$, exists over the entire crystal, and therefore " sees" the whole crystal.

5.3 KRONIG–PENNEY MODEL OF BAND STRUCTURE

The potential depicted in Fig. 5.6 is fairly realistic in a one-dimensional sense, although Schrödinger's equation cannot be solved exactly when V has this form. As an approximate model, assume a one-dimensional crystal where

$$V(x) = \begin{cases} 0, & 0 \le x \le a_1, \\ V_0, & -a_2 \le x \le 0, \end{cases}$$

and where $a = a_1 + a_2$ is the period of the lattice, as shown in Fig. 5.7. This is known as the *Kronig–Penney model.*[†]

[†] This model provides an introduction to band theory. However, it is overly simplistic and leads to only a very rough approximation of carrier behavior in real materials.

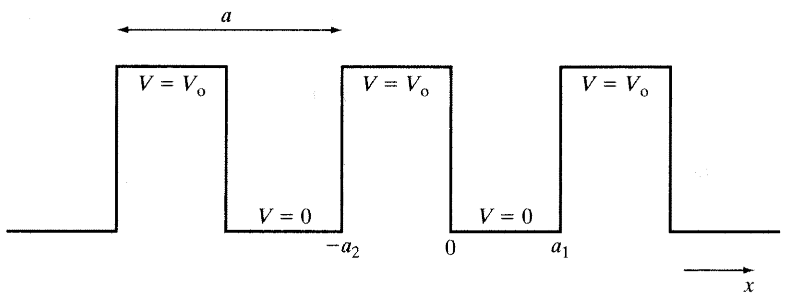


Figure 5.7 Kronig–Penney model of the potential due to a periodic lattice of period a.

In the region $-a_2 \leq x \leq 0$, the potential is $V = V_0$, and the solution of Schrödinger's equation is

$$\psi(x) = Ae^{i\beta x} + Be^{-i\beta x}, \tag{5.10}$$

where

$$\beta = \sqrt{\frac{2m_e(E - V_0)}{\hbar^2}}, \tag{5.11}$$

as shown in Section 4.1.1. In a similar manner, for $0 \leq x \leq a_1$ the potential is $V = 0$, and the solution of Schrödinger's equation is

$$\psi(x) = De^{i\alpha x} + Fe^{-i\alpha x}, \tag{5.12}$$

where

$$\alpha = \sqrt{\frac{2m_e E}{\hbar^2}}. \tag{5.13}$$

In one dimension, the Bloch form (5.7) is

$$\psi(x) = u(x)\, e^{ikx}, \tag{5.14}$$

in terms of the to-be-determined Bloch wavevector k, where

$$u(x) = u(x + a). \tag{5.15}$$

Therefore,

$$\begin{aligned} \psi(x \pm a) &= u(x \pm a)\, e^{ik(x \pm a)} = u(x)\, e^{ik(x \pm a)} \\ &= \psi(x)\, e^{\pm ika}, \end{aligned} \tag{5.16}$$

and, using this relationship, we can write down the wavefunction in the following period, $a_1 \leq x \leq a_1 + a$, as

$$\begin{aligned} \psi(x) &= \left\{Ae^{i\beta(x-a)} + Be^{-i\beta(x-a)}\right\} e^{ika}, \qquad a_1 \leq x \leq a, \\ &= \left\{De^{i\alpha(x-a)} + Fe^{-i\alpha(x-a)}\right\} e^{ika}, \qquad a \leq x \leq a_1 + a. \end{aligned} \tag{5.17}$$

Enforcing continuity of ψ and ψ' at $x = 0$ and $x = a_1$ leads to the eigenvalue equation

$$\cos ka = \cos(\alpha a_1)\cosh(\delta a_2) - \frac{\alpha^2 - \delta^2}{2\alpha\delta}\sin(\alpha a_1)\sinh(\delta a_2), \tag{5.18}$$

if $0 < E < V_0$, and

$$\cos ka = \cos(\alpha a_1)\cos(\beta a_2) - \frac{\alpha^2 + \beta^2}{2\alpha\beta}\sin(\alpha a_1)\sin(\beta a_2), \tag{5.19}$$

if $E > V_0$, where

$$\delta = \sqrt{\frac{2m_e(V_0 - E)}{\hbar^2}}. \tag{5.20}$$

In the preceding equations, the energy E is the only unknown parameter. For a solution to exist, we must have

$$-1 \leq \cos ka \leq +1, \tag{5.21}$$

and so the right side of (5.18), denoted as $r(E)$, must obey this condition. A typical plot of $r(E)$ vs. E is shown in Fig. 5.8.

This figure makes clear the fact that there are certain allowed values of energy, called *allowed energy bands*, and certain unallowed values of energy, called *band gaps*. That is, if E is in an allowed energy band, Schrödinger's equation for the Kronig–Penney model has a solution, and if E is not in an allowed energy band, there is no solution (because we

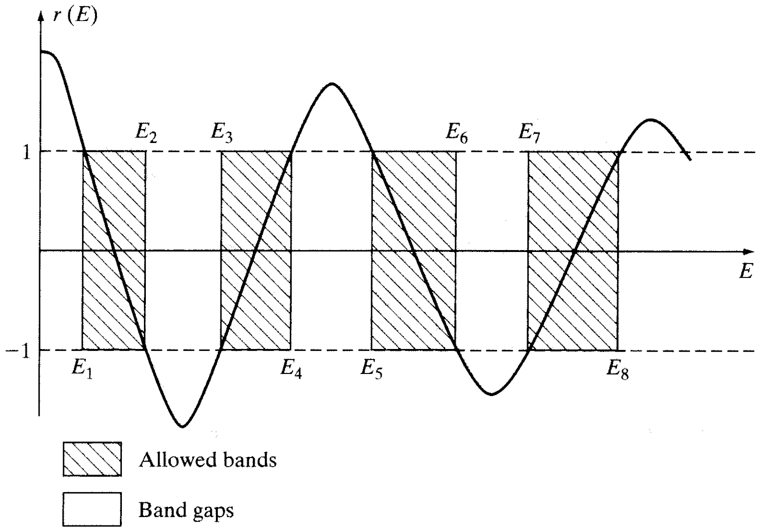


Figure 5.8 Plot of $r(E)$ versus E, showing allowed energy bands and band gaps.

must have $|\cos ka| \leq 1$ in (5.21)). Within an allowed band energy can take any value, i.e., it is not discretized. Note that as energy increases, the allowed energy bands increase in width, and so the forbidden bands decrease in width. As we will see later in this chapter, the concept of energy bands leads to the fundamental characteristics of conductors, insulators, and semiconductors.

Since

$$\cos ka = r(E), \tag{5.22}$$

we can generate an important figure called the *dispersion diagram*, which is a diagram of energy versus wavenumber (E vs. k). To generate the dispersion diagram, start at $E = 0$ and compute $r(E)$. If $|r(E)| = |\cos ka| > 1$, we are at a forbidden energy (i.e., in a bandgap), and we need to increase E a bit and try again. If $|r(E)| \leq 1$, we are at an allowed energy (i.e., in an energy band), and in this case the corresponding wavenumber is

$$k = \frac{1}{a}\cos^{-1}(r(E)). \tag{5.23}$$

Since cosine is an even function, $-k$ will also be a solution. By increasing E by a small amount and checking the value of $r(E)$, we can generate the plot of allowed and unallowed energy bands. One form of the result will look like Fig. 5.9. This depiction is known as the *extended zone scheme.*

The various sections of wavenumber space are divided into what are called *Brillouin zones*, with the range

$$-\frac{\pi}{a} \leq k \leq \frac{\pi}{a} \tag{5.24}$$

denoting the important *first Brillouin zone*. The second Brillouin zone is the range

$$-\frac{2\pi}{a} \leq k < -\frac{\pi}{a}, \quad \frac{\pi}{a} < k \leq \frac{2\pi}{a}, \tag{5.25}$$

and so on for higher zones.

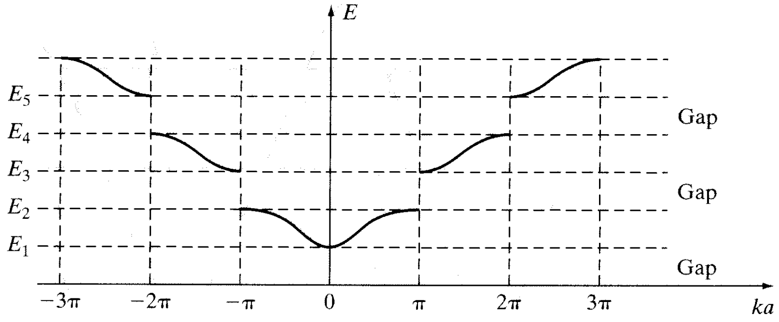


Figure 5.9 Energy band energy E versus Bloch wavenumber k in the extended zone scheme.

Due to the periodic nature of the crystal, there is no real difference between the wavevector k and the wavevector $k \pm n2\pi/a$, $n = 0, \pm 1, \pm 2, \ldots$ (that is, the structure is periodic in both "real space" and "wavevector space," also known as *reciprocal space*). This can be proved in general, although here one can simply note that (5.23) is multivalued due to the inverse cosine function. That is, assume that a given E value determines the right side of (5.18) (or (5.19)), leading to $\pm k$ from (5.23). Any other wavenumber value $\pm k + n2\pi/a$, $n = 0, \pm 1, \pm 2, \ldots$ will also satisfy (5.18) (or (5.19)), since $\cos(k + n2\pi/a)a = \cos ka$ for n an integer. Therefore, for a given allowed energy, there are an infinite number of k values, and the E–k plot can also be drawn with E oscillating in a continuous fashion within the band as k increases. This depiction, shown in Fig. 5.10, is known as the *repeated zone scheme*).

Lastly, another depiction arises from noting that the energy bands in the higher Brillouin zones can be all translated to the first Brillouin zone by shifts of $n2\pi/a$. This results in what is called the *reduced zone scheme*, depicted in Fig. 5.11. In the remainder of the text, band diagrams will be depicted in the reduced zone scheme, which is the most common format for describing band structure. However, it should be noted that all three depictions (extended, repeated, and reduced) convey the same information.

5.3.1 Effective Mass

In the previous chapter, various nanostructures, such as quantum wells, quantum wires, and quantum dots, were considered. However, the structures were basically empty boxes for electrons. At this point, it may seem difficult to imagine how one can incorporate the fact that a nanostructure is made from a real material, such as a semiconductor. It turns out that there is a relatively simple approximate method to do this, involving a concept known as *effective mass*.

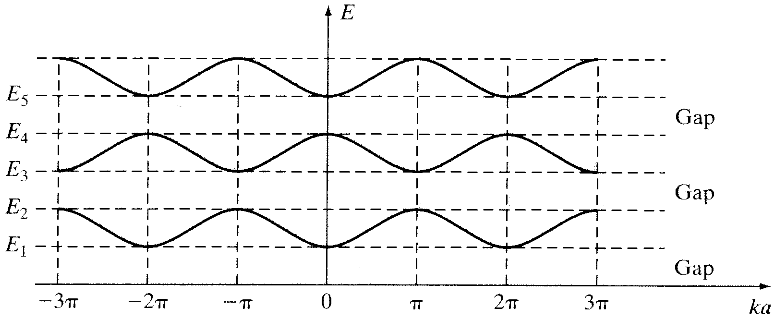


Figure 5.10 Energy band energy E versus Bloch wavenumber k in the repeated zone scheme.

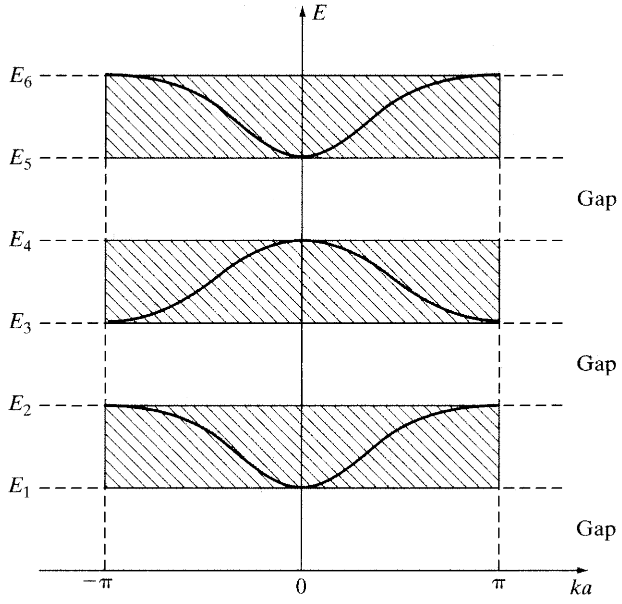


Figure 5.11 Energy band energy E versus Bloch wavenumber k in the reduced zone scheme.

In a nonrelativistic sense, an electron in empty space has a well-defined, constant mass. However, it is often useful to view mass as merely a proportionality factor between force and acceleration. (Recall Newton's second law, $\mathbf{F} = m\mathbf{a}$.) This view is particularly appropriate when studying electrons in crystals, since in this case it turns out that often electrons appear to act as if their mass is different from the free-space value. This simply means that in a crystal, electrons do not respond to external forces in the same way that free electrons do.

We view an electron quantum mechanically as represented by a wavepacket, with the electron's velocity being its group velocity (as discussed in Section 2.5),

$$v_g = \frac{\partial \omega}{\partial k} = \frac{1}{\hbar}\frac{\partial E}{\partial k}. \tag{5.26}$$

The influence of the electron's environment is contained in the energy relation $E(k)$. For example, for an electron in free space, we use (4.5),

$$E(k) = V_0 + \frac{\hbar^2 k^2}{2m_e}. \tag{5.27}$$

For an electron in a periodic potential, even assuming the simplistic Kronig–Penney model, $E(k)$ cannot be given by a simple formula, but it can be determined by the procedure developed on pages 140. Methods for determining the energy-wavenumber relationship for more realistic models of materials are beyond the scope of this book, although the band structure can be exceedingly complex. However, usually the complete band structure of a material does not need to be known, since often only electrons in certain regions of a band are of interest. For example, considering semiconductors, typically one is interested in the behavior of electrons near a band edge (say, in the center of the first Brillouin zone near $k = 0$), since these are the electrons that will be most important for conduction. In that case, only the local behavior of the E–k curve will be important.

The main idea is the following. We know how to solve Schrödinger's equation for an electron in an "empty" region of space, either confined or unbounded, where we obtain a parabolic E–k relationship. (See, e.g., (4.5) or (4.26).) The parabola has the form $E = V_0 + \alpha k^2$, where $\alpha = \hbar^2/2m_e$. Considering the importance of electrons near bandedges in semiconductors, and the occurrence of parabolic-like dispersion behavior near bandedges in real materials (the band structure of Ge, Si, and GaAs is shown in Fig. 5.12), it makes sense

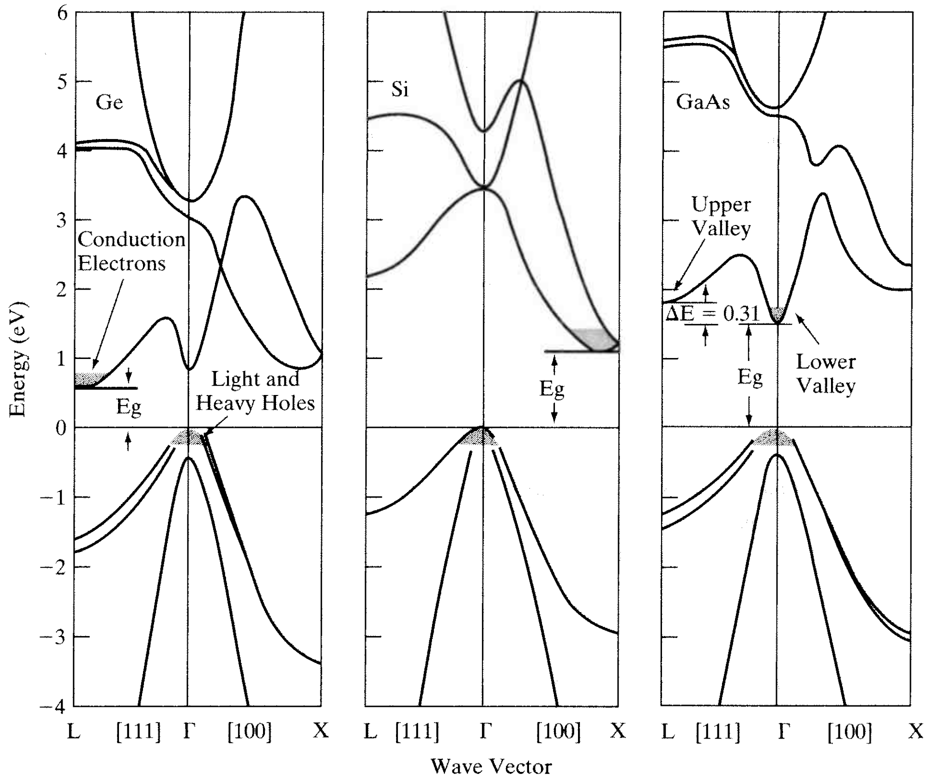


Figure 5.12 Bandstructure of Ge, Si, and GaAs. The Γ point is in the center of the first Brillouin zone ($k = (0, 0, 0)$), and the X and L points represent the zone boundaries along certain directions in three-dimensional k−space. In this case, $X = (1, 0, 0)\, 2\pi/a$, and $L = (1, 1, 1)\, \pi/a$, where a is the length of the cube edge in the fcc lattice. The curves do not look symmetric about the center point since moving from Γ to X and from Γ to L represents moving in different physical directions in the crystal. (Based on a figure from S. M. Sze, *Physics of Semiconductor Devices*, John Wiley & Sons, 1969. Used by permission.)

to locally model dispersion behavior using a parabolic relationship. This leads to the form

$$E(k) = E(k_0) + \beta (k - k_0)^2, \tag{5.28}$$

which is expected to be valid near bandedges centered at the point $k = k_0$, and where β will depend on the material in question (i.e., β governs the rate of expansion of the parabola). Since the band structure can be calculated or measured, β can be determined for a given material.[†] However, (5.28) can be made to look like the "empty" space result if we set $\beta = \hbar^2/2m^*$, where m^* is called the effective mass (which will, like β, depend on the material and band in question). Substituting this value of mass for m_e in Schrödinger's equation essentially incorporates the effect of the material on the electron, as long as one is interested in the behavior of the electron in the vicinity of a parabolic region of a band.

An explicit equation for m^* can be obtained by taking two derivatives of (5.28) with respect to k, leading to

$$m^* = \hbar^2 \left(\frac{\partial^2 E}{\partial k^2} \right)^{-1}. \tag{5.29}$$

Note that the effective mass is proportional to the reciprocal of the curvature of the E–k plot. From (5.26),

$$\frac{\partial v_g}{\partial k} = \frac{1}{\hbar} \frac{\partial^2 E}{\partial k^2}, \tag{5.30}$$

and so an equivalent expression for effective mass is

$$m^* = \hbar \left(\frac{\partial v_g}{\partial k} \right)^{-1}. \tag{5.31}$$

One can also develop a different definition of effective mass. Using $\hbar k$ as momentum,[‡] with an electron represented by a wavepacket moving at the group velocity, one can define an effective mass such that

$$m^* v_g = \hbar k. \tag{5.32}$$

This is consistent with Newton's law (force equals the time rate of change of momentum),

$$F = \frac{dp}{dt}, \tag{5.33}$$

[†] As seen in Fig. 5.12, for any material there will be many different regions of parabolic E–k behavior, and, thus, many different values of β will exist, although typically only one or a few such parabolic regions will be of interest.

[‡] The quantity $\hbar k$ is called the *crystal momentum*. This is *not* the physical momentum of the electron. Rather, it is a quantity that describes the electron's Bloch state within a band, and is the correct quantity to use in Newton's law to obtain the electron's dynamics in a crystal.

leading to the equation of motion

$$q_e \mathcal{E} = \hbar \frac{dk}{dt}, \tag{5.34}$$

where $\mathcal{E}$ is the magnitude of the electric field for this scalar problem. Using (5.26) in (5.32), we obtain

$$m^* = \hbar^2 k \left(\frac{\partial E}{\partial k} \right)^{-1}. \tag{5.35}$$

As discussed in [8], this definition of effective mass can be used where the E–k curve is not parabolic. If the E–k relation is simply $E = V_0 + \beta k^2$, the two definitions lead to the same effective mass, $m^* = \hbar^2/2\beta$. The important point is that at most points of interest within an allowed energy band, the electron in a crystal moves as if it were free, except with an effective mass.

Effective mass can be positive or negative. A positive effective mass $m^*/m_e > 1$ means that the electron's velocity increase is less than what it would be for an electron in free space (i.e., the electron seems heavier), the difference indicating momentum transfer to the lattice. A positive effective mass $m^*/m_e < 1$, the usual case in semiconductors at the bandedge, means that the electron's velocity increase is more than it would be for an electron in free space (i.e., the electron seems lighter), the difference indicating momentum transfer from the lattice. For example, for GaAs, $m_e^* = 0.067 m_e$, resulting in a significant narrowing of the E–k curve, as depicted in Fig. 5.13. A negative effective mass (often

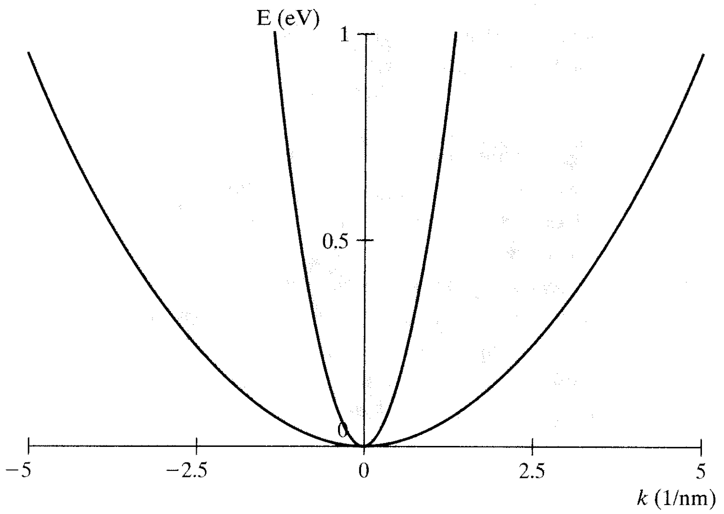
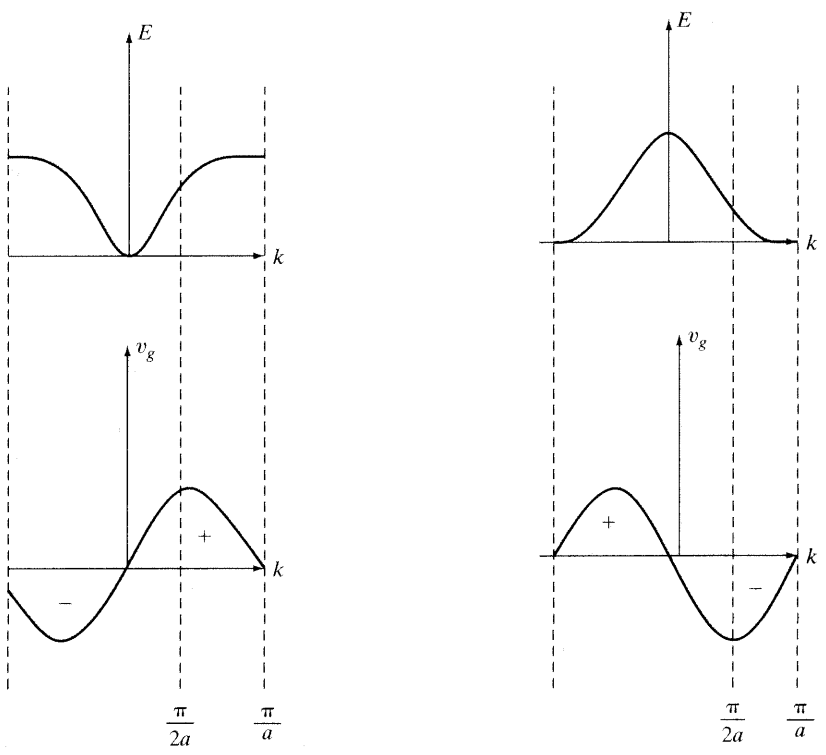


Figure 5.13 E (eV) versus k (nm^{-1}) for an electron in free space, $E(k) = \hbar^2 k^2/2m_e$ (wider curve), and for an electron at the bottom of the conduction band (set to reference level $E = 0$) in GaAs using the parabolic approximation, $E(k) = \hbar^2 k^2/2m_e^*$ (narrower curve). The effective mass $m_e^* = 0.067 m_e$ for GaAs results in significant narrowing of the E–k curve relative to the free-space environment.



(a) (b)

Figure 5.14 Energy and group velocity $v_g = (1/\hbar)(\partial E/\partial k)$ for two representative bands (left, a typical conduction band, and right, a typical valence band; bands will be discussed further in the next section). The period of the crystal is a.

found near the top of an energy band) means that, conversely, the particle is accelerated in a direction opposite to the direction of the applied force.[†]

Figure 5.14 depicts the relationship between energy and group velocity for a typical portion of an E–k curve.

If we consider what happens when an electric field is applied to a crystal, a clearer picture of the behavior of electrons in energy bands develops. When a constant electric field of magnitude $\mathcal{E}$ is applied to the crystal, an electron will experience a Lorentz force, $q_e\mathcal{E}$. This force will accelerate the electron, increasing its wavenumber. To be specific, the equation of motion (5.34) can easily be solved to yield

$$k(t) = k(0) + \frac{q_e\mathcal{E}}{\hbar} t, \tag{5.36}$$

[†] To see this, consider $\mathbf{F} = m^*\mathbf{a}$. Since $\mathbf{F} = q_e\mathbf{E}$ with $q_e < 0$, then $m^* < 0$ means that in response to an applied electric field, the electron moves in a direction opposite to that which would occur in free space.

and so k increases linearly with respect to time. (For simplicity, we can assume $k(0) = 0$.) As k increases towards the value $k = \pi/2a$, the electron's velocity increases, as expected, as does the effective mass.

At $k = \pi/2a$, the E–k curve has an inflection point (and so the parabolic band assumption is no longer valid), and the electron's velocity has reached a maximum. As k increases further, the electron decelerates, and finally, as k reaches the Brillouin zone boundary at π/a, the electron's velocity goes to zero, indicating that the electron wavefunction is represented by a standing, rather than traveling, wave.

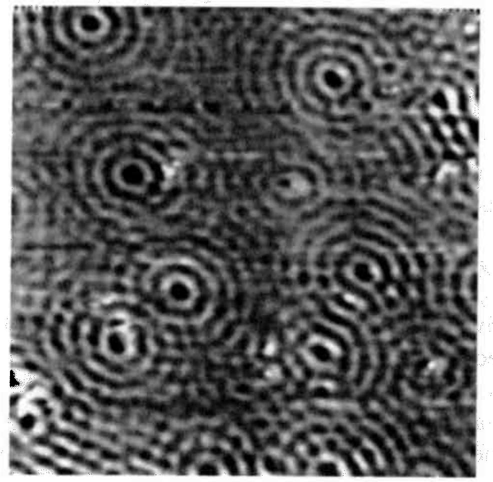
As k increases further and moves away from the zone boundary into the second Brillouin zone,[†] a conduction band electron will initially have a negative group velocity, indicating that the electron is moving in the reverse direction. The electron continues to reverse direction periodically as k increases through zone boundaries.

Therefore, we have the following picture. We apply a static (d.c.) electric field to the crystal, and as a result of band structure, the electron oscillates back and forth. (These are called *Bloch oscillations*.) Can this really be so? In practice, this behavior is not seen. Defects and impurities will generally be present,[‡] as well as lattice vibrations called *phonons*,[§] and

---

[†] Despite what could be inferred from the extended zone scheme depicted in Fig. 5.9, as k increases through Brillouin zone boundaries, the energy of the electron does not "jump." $E(k)$ evolves smoothly, as depicted in the repeated zone scheme of Fig. 5.10. Sometimes the electron is said to be Bragg scattered at the zone boundaries, but this gives a rather unphysical picture, and so this idea is avoided here.

[‡] As an interesting aside, the following figure shows an STM image of electron standing waves extending outward from defects/impurities in a Cu surface.



A 42 nm × 42 nm STM image of Cu ($T = 150$ K). Defects create standing electron waves emanating radially outward from the defect. (Reprinted with permission from Petersen, L., P. Laitenberger, E. Lægsgaard, and F. Besenbacher, "Screening Waves from Steps and Defects on Cu(111) and Au(111) Imagined with STM: Contribution from Bulk Electrons," *Phys. Rev. B*, 58 (1998): 7361. Copyright 1998, American Physical Society.)

[§] Phonons are lattice vibrations (vibrations of the atoms in a lattice) treated quantum mechanically. The energy of these vibrations (i.e., heat) is related to the kinetic energy of the atoms in the material. For a lattice vibration of frequency ω, energy is quantized as $\hbar\omega$, as with all quantum particles. Heat is carried through a the electron will collide with these obstacles such that the wavevector will not be able to increase very much. That is, we can only solve (5.34) over the time period between collisions, and, therefore, the wavevector does not simply linearly increase. This is discussed further in Section 10.1. The wavevector only has a chance to increase a small amount in the first zone before a collision occurs, and, therefore, the k value will not reach the zone boundary $k = \pm\pi/a$.

For example, assume that the average time between collisions (called the *momentum relaxation time*) is $\tau = 2.47 \times 10^{-14}$ s, which is the value for copper at room temperature. If $\mathcal{E} = -100$ V/m, then $k\left(t = 2.47 \times 10^{-14}\right) = 3{,}753$ m^{-1}. The first Brillouin zone boundary occurs at $k = \pm\pi/a$, and, using $a = 3.61$ Å for the lattice constant for copper, we obtain $k = 8.7 \times 10^{9}$ m^{-1} at the zone boundary. Thus, for a typical collision time, the wave vector only gets about 10^{-5} percent of the way to the Brillouin zone boundary before its momentum is stopped by a collision! It is obvious that even very large applied fields or long values[†] of τ will not change the situation too much. However, Bloch oscillations are real, and have been experimentally observed in a number of periodic structures, most notably semiconductor superlattices. (See Section 6.3.4.)

Note that electron collisions are regarded as the *cause* of electrical resistance in the classical, free-electron gas model of conductivity, as discussed in Section 10.1. However, in the quantum mechanical picture, the collisions are actually *necessary* to allow d.c. current to flow! If no collisions occur, such as may be the case in an ultrapure sample at low temperatures, an a.c. current will result from a d.c. applied field!

Lastly, note that for free electrons, the E–k curve (4.5) is a parabola,

$$E = \frac{\hbar^2 k^2}{2m_e}, \tag{5.37}$$

and, therefore,

$$\frac{\partial^2 E}{\partial k^2} = \frac{\hbar^2}{m_e}, \tag{5.38}$$

such that

$$m^* = m_e, \tag{5.39}$$

i.e., the effective mass is the ordinary mass, as expected. It is because the dispersion relation for an electron in a material does not have the simple form (5.37) that the concept of effective mass arises.

In summary, via the effective mass, we represent the influence of a crystalline lattice on an electron. The specific nature of the periodic potential is contained in the $E\,(k)$ relationship, which usually must be obtained numerically or experimentally. We can often greatly simplify solving problems involving an electron in a crystal by solving an equivalent problem of an electron in free space, except where the mass of the electron is given by m^*. For example, we can solve Schrödinger equation

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V\,(\mathbf{r})\right)\psi\,(\mathbf{r}) = E\psi\,(\mathbf{r}), \tag{5.40}$$

where $V\,(\mathbf{r})$ accounts for the crystal lattice (and is, therefore, horrendously complicated), or we can solve the much simpler equation

$$\left(-\frac{\hbar^2}{2m^*}\nabla^2 + V_0\right)\psi\,(\mathbf{r}) = E\psi\,(\mathbf{r}), \tag{5.41}$$

where m^* is the effective mass, accounting for the crystalline lattice, and where V_0 is a constant potential depending on the problem. The method of replacing (5.40) with (5.41) is known as the *effective mass approximation* of Schrödinger's equation.

The effective masses of electrons, m_e^*, and holes, m_h^*, in Si and GaAs are given in Table 5.1, and values for some other common semiconductors are given in Table IV in Appendix B. Holes will be discussed later in this chapter.

To further complicate the situation, in many crystals there is a different effective mass in each different direction; near bandedges this is given as

$$m^*_{x,y,z} = \hbar^2 \left(\frac{\partial^2 E}{\partial k^2_{x,y,z}}\right)^{-1}. \tag{5.42}$$

However, for many metals, the effect of the lattice is generally screened by the large density of electrons (Section 3.5), such that often in a metal one can set $m^* = m_e$.

In an amorphous material such as SiO_2, the concept of energy bands is not quite appropriate. However, often electrons in SiO_2 can be modeled as having an effective mass of the order of $0.4m_e$–$0.9m_e$, and the effective mass Schrödinger equation can be used to model electrons in, for example, metal-oxide-semiconductor structures such MOS capacitors.

TABLE 5.1 EFFECTIVE MASS IN Si AND GaAs.

	Effective Mass	
Semiconductor	m_e^*/m_e	m_h^*/m_e
Si	0.26	0.50 0.24
GaAs	0.067	0.50 0.082

---

material as a flow of phonons. Similar to photons, phonons transport energy and momentum, but not mass, and obey Bose–Einstein statistics. (See Section 8.2.)

There are two broad classes of phonons: relatively low-energy modes called acoustic phonons, which are related to sound propagation in a material, and higher energy modes called optical phonons, so named because optical energies can excite them.

[†] τ is temperature and material dependent, although one could consider practical values into the picosecond range.

5.4 BAND THEORY OF SOLIDS

The material presented in the previous section leads to what is called the *band theory of solids*, the central idea of which is that, due to the periodic potential associated with the crystalline lattice, there are allowed and disallowed energy bands. Furthermore, we have only found the possible bands and the bandgaps, but have not considered if a certain band will be "filled." Whether or not a band has electrons in it depends on the number of electrons in the system, and the energy of the electrons. Several different situations arise:

1. If an allowed band is completely empty of electrons, obviously there are no electrons in the band to participate in electrical conduction. This can happen, for example, in a high-energy band, where the energies of the band are above the energies of any of the system's electrons.
2. Perhaps a bit more surprising is the fact that when an allowed band is completely filled with electrons, those electrons cannot contribute to electric conduction either. This is because electrons are fermions, and must obey the Pauli exclusion principle, so that no two electrons can be in the same state. Therefore, if an electron is given energy, due to, say, a voltage applied to the material, the electron must move to a (perhaps only slightly) higher energy state. However, if all such states are already filled, the electron has no empty state to move into, and, therefore, the electron cannot gain any energy and contribute to conduction.[†] This would be the case, for example, for a band of energies that are much less than the Fermi energy of the system's electrons.

To partially summarize, no conduction can take place in a material that has energy bands that are either completely filled or completely empty. An analogy is to a jar of marbles. An empty band is like an empty jar, and a filled band is like a jar filled so full that no marbles can move (thus, there are no empty spots for a marble to move into). Only when the jar is partially full can marbles move within the jar.

3. Thus, we are left to conclude that only electrons in a partially filled energy band can contribute to conduction.

In general, the lower bands in a material will be completely filled with electrons. (Recall that in Section 4.6 we discussed the hydrogen atom and the periodic table, and that states are typically filled one by one, starting at low energies.) The main question is whether or not the uppermost band that contains electrons is completely filled, or only partially filled. Materials that have partially filled uppermost bands are called conductors (mostly metals), such that when energy is supplied by an external source, electrons can move into a higher unoccupied state within the band.

To illustrate these concepts, a *real-space band diagram* for a typical metal is shown in Fig. 5.15. It is obvious that the highest band to contain electrons is partially full, and so electrical conduction can take place.

[†] Of course, if a very large amount of energy is supplied such that the electron can jump across the energy gap into a different, higher, empty, or partially filled band, that electron can contribute to conduction.

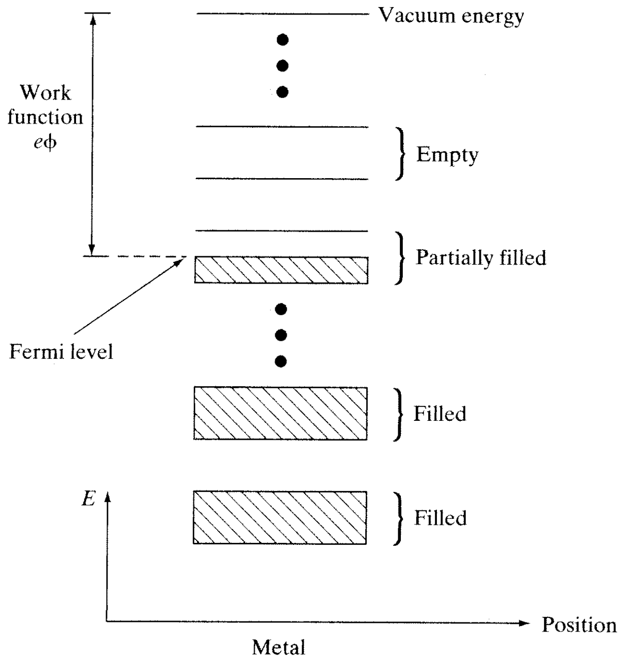


Figure 5.15 Band structure (real space) for a typical metal.

The *work function*, $e\phi$, is the energy difference between the vacuum level and the Fermi level (the uppermost energy of the electrons at $T = 0$ K). It is thus the energy needed to completely liberate an electron from the metal at $T = 0$ K, and, given the narrow tail of the Fermi–Dirac distribution, which will be described in Section 8.2, this energy approximately holds at other temperatures as well. Energy can be supplied, for example, thermally, or electromagnetically, as in the photoelectric effect. For example, considering shining light on a metal surface, *Einstein's relation* is

$$\hbar\omega = e\phi + E_{KE}, \tag{5.43}$$

where $\hbar\omega$ is the energy of the incident light, and E_{KE} is the kinetic energy of the emitted electrons. For copper, the work function is on the order of 4.75 eV (ranging from 4.5 to 5.0 eV depending on the crystal orientation), and for gold, $e\phi$ is in a similar range. Thus, as discussed in Section 2.2, the photoemitted electrons will have kinetic energy $E_{KE} = \hbar\omega - e\phi$, which only depends on frequency, and not on intensity.

Materials having completely filled lower bands and empty upper bands, such that conduction cannot take place, are either insulators or semiconductors at low temperatures, as shown in Fig. 5.16. The highest band that is filled at $T = 0$ K in semiconductors is called the *valence band*, and the unfilled bands above the valence band are called *conduction bands*.

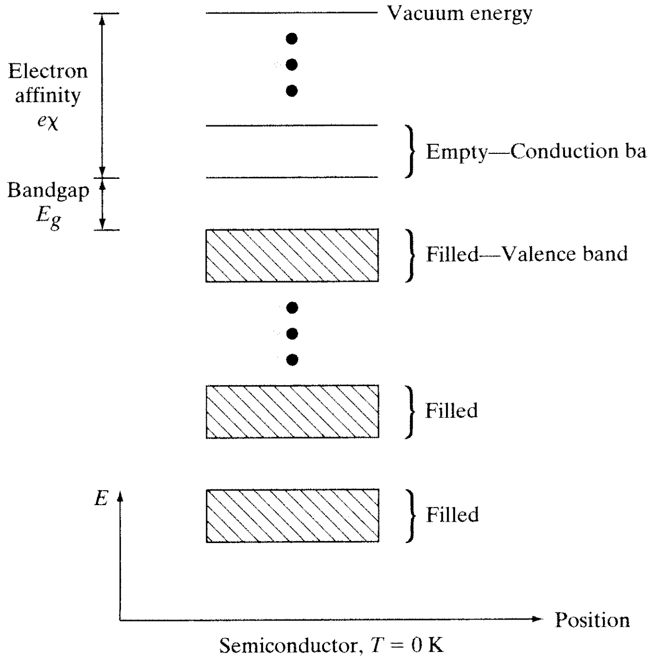


Figure 5.16 Band structure for a typical semiconductor ($E_g \sim 1$ eV) at $T = 0$ K. This picture also holds for a typical insulator ($E_g \sim 8$ eV) for a range of temperatures.

From the point of band theory, the main difference between a semiconductor and an insulator is the size of the bandgap. For semiconductors, the energy gap is not too large (for Si, $E_g \simeq 1.12$ eV, and for GaAs, $E_g \simeq 1.43$ eV), so that as temperature increases some thermally excited electrons have enough energy to jump across the gap and reach a previously empty band, thus being able to contribute to conduction. Other than thermal energy, an external excitation such as an applied voltage can also supply enough energy for electrons to jump across the gap and contribute to conduction, especially when dopant atoms are present in semiconductors (to be discussed in the next section). The band structure after some energy input (in this case, thermal) is depicted in Fig. 5.17. The energy difference between the vacuum level and the bottom of the conduction band is called the *electron affinity*, $e\chi$.

For insulators, the bandgap is large (perhaps 8–10 eV; SiO_2 has $E_g \simeq 8$ eV), so that thermally excited electrons, or those excited by an applied voltage, do not generally have enough energy to cross the gap. Of course, even for good insulators, the actual gap is finite. If enough energy is supplied, thermally or by an applied voltage, electrons will cross the gap and move into an unoccupied state in a higher band, thus being able to contribute to conduction.

For example, consider the gate oxide in a MOSFET assuming a $V_0 = 1.5$ volt power supply and SiO_2 as the gate oxide. The associated energy is $eV_0 = 1.5$ eV, which is much less than the bandgap energy of SiO_2. Therefore, in this case, valence electrons will not be able to cross the gap and contribute to conduction. However, this energy may be large enough to excite electrons associated with impurities, (which lie at energy levels in the band gap, as discussed in the next section), into the conduction band.[†]

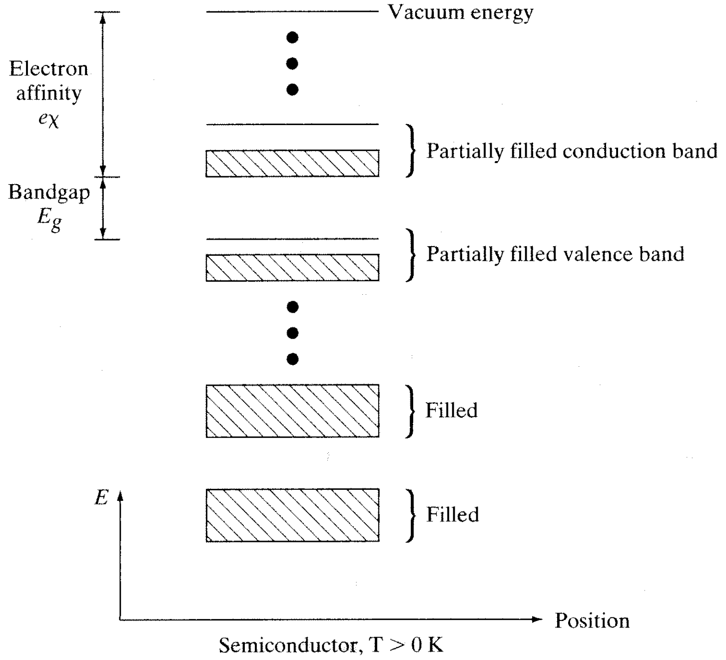


Figure 5.17 Band structure for a typical semiconductor at $T > 0$ K, showing that some of the electrons that were in the valence band at $T = 0$ K (Fig. 5.16) have moved, via thermal energy, into the first conduction band. Note that E_g is slightly temperature dependent.

Recalling the electronic configuration of the elements in the periodic table, we find that if the outermost shell, (which, for an individual atom, is analogous to the topmost conduction band containing electrons for a solid), is partially filled, due to an odd number of electrons, the material will generally be a conductor. If the outermost shell is completely filled, due to an even number of electrons, the material will be an insulator or a semiconductor, depending on the size of the bandgap. This is sort of a general rule of thumb, and doesn't always hold. For example, beryllium has four electrons, yet is a metal. (Beryllium belongs to the class of divalent metals in group IIA of the periodic table.)

Holes. When an electron is elevated from the valence band to a conduction band, it leaves an empty spot in the valence band called a *hole*. Holes are not implicated in an electron gas model of metals, nor generally in the characterization of metals at all, due to the very large numbers of "free" conduction band electrons. However, in a semiconductor, holes play a prominent role, and for each electron that contributes to conduction, there is an associated hole that can also contribute to conduction. It can be seen that holes are capable

---

[†] We reiterate that SiO_2 is an amorphous material, and, as such, the concept of a bandgap, which was developed for crystals, is an approximation.

of moving in the valence band by considering the following. Assume that electron e_1 moves to the conduction band, leaving behind a hole h_1 in the valence band. The valence band is no longer full, and, therefore, another valence band electron e_2 can move, and it will tend to move into h_1. This fills the original hole, but leaves another hole, h_2, in the position that e_2 previously occupied. A valence band electron e_3 moves into h_2, leaving behind hole h_3, and so on. In this way, holes (spots in the lattice where an electron is missing) can move in the valence band. Holes carry positive charge (i.e., the absence of negative charge), and move in the direction of an applied electric field. Physically, it should be kept in mind that in both bands it is actually electrons that are moving.

Furthermore, the hole effective mass in many semiconductors is separated into heavy and light hole masses, since the valence band is often divided into light and heavy bands. (See Fig. 5.12 on page 143.) Actually, the valence band often consists of several bands, and in each band the effective mass is different. Bands having large effective mass are called *heavy bands*, and bands having lower effective mass are called *light bands*. The heavy and light hole effective masses for Si and GaAs are listed in Table 5.1, and values for other common semiconductors are provided in Table IV in Appendix B.

5.4.1 Doping in Semiconductors

It is often necessary to know how many electrons can contribute to conduction. In metals, the conduction band is partially full, even at $T = 0$ K, and there are a lot of electrons to contribute to electrical conduction.[†] As described previously, in semiconductors at $T = 0$ K, the valence band is full and the conduction band is empty, so that no electrons can participate in conduction. As temperature increases, thermal energy of the crystal elevates some electrons into the conduction band, leaving a hole in the valence band. The resulting (*intrinsic*) carrier densities of electrons and holes are labeled n_i and p_i, respectively. It is obvious that, since each time an electron is elevated into the conduction band a hole is created in the valence band, $n_i = p_i$.

The intrinsic carrier concentration in a typical semiconductor is very low because thermal energy is usually small compared to the bandgap energy. As a result, in Si at room temperature, for example, $n_i \sim 1.5 \times 10^{10}$ cm^{-3}; for comparison, copper has on the order of 10^{22} cm^{-3} carriers. However, the carrier concentration in semiconductors can be modified by adding impurities called *dopants*. Undoped semiconductors are called *intrinsic semiconductors*, and doped semiconductors are called *extrinsic semiconductors*.

For illustration purposes, consider Si, a Group IV element (i.e., silicon has four outer shell electrons). If we introduce impurity elements from Group V (i.e., an element having five outer shell electrons, such as phosphorus), each impurity will bond (covalently) with four neighboring silicon atoms, using up four of its electrons in the process, as shown in Fig. 5.18. The fifth electron will still be bound to the nucleus of the impurity (i.e., the impurity atom will have a net positive charge), although much more weakly than it would be in the elemental state. A small amount of energy (typically, on the order of 0.05 eV) will

---

[†] That there are *some* electrons to contribute to conduction in a partially filled band is obvious. That there are, in fact, *a lot* of electrons that can contribute to conduction is established by considering the density of states, which will be discussed in Chapter 8.

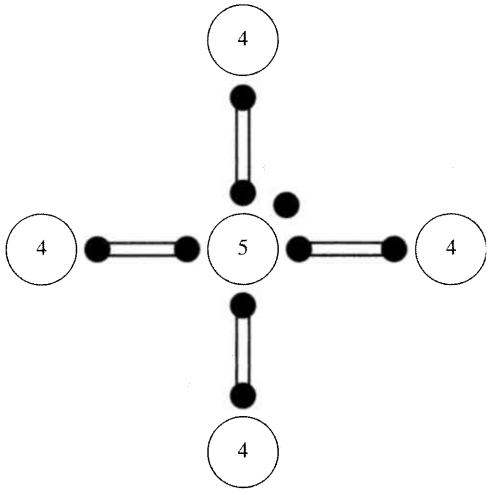


Figure 5.18 Part of a Group IV element lattice with the insertion of a Group V element. The double parallel lines indicate a covalent bond. One "extra" electron will be present, loosely bound to the Group V atom.

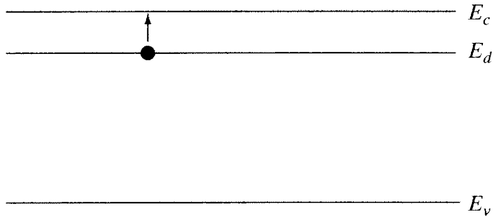


Figure 5.19 Real-space energy band diagram showing the level of a donor electron, and its transition to the conduction band upon adding a small amount of energy.

liberate the electron and move it to the conduction band. The impurity is called a *donor*, and we can view this on an energy band diagram as the electron being at a donor level, E_d, lying just below the conduction bandedge, as shown in Fig. 5.19. A small amount of thermal energy will be able to bridge the energy gap $E_c - E_d$, and the electron will contribute to conduction. For example, considering silicon with the donor being phosphorus, $E_c - E_d \simeq 0.044$ eV at room temperature. Thus, donor atoms are easily ionized, resulting in what is called an *n type* semiconductor.

As a rough approximation, since the excess electron is loosely held to the donor atom, the situation resembles the hydrogen atom in that one can consider a two-particle problem (the excess electron, and everything else). This hydrogen model can provide a rough approximation for these shallow dopants. The fact that the dopant atom is immersed in a semiconductor, rather than free space, is accounted for by using the appropriate effective mass and permittivity. Energy states of the hydrogen atom are given by (4.117), and substituting the effective mass in place of m_e, and the permittivity of the semiconductor in place of the vacuum permittivity, we have

$$E_n = -\frac{m_e^* q_e^4}{8\varepsilon_r^2 \varepsilon_0^2 h^2 n^2}. \tag{5.44}$$

If this dopant atom is to be ionized, there must be a transition from the $n = 1$ state to $n \rightarrow \infty$, leading to the donor ionization energy

$$\Delta E_d = \frac{m_e^* q_e^4}{8\varepsilon_r^2 \varepsilon_0^2 h^2}. \tag{5.45}$$

For Si, using $m_e^* = 0.26 m_e$ and $\varepsilon_r = 11.7$, we obtain $\Delta E_d = 0.025$ eV, within an order of magnitude of measured values. Measured values of donor ionization energies for Si and Ge are given in Table 5.2.

In a similar manner, the impurity element can come from Group III of the periodic table, having three outer shell electrons (e.g., boron). In forming a covalent bond with its four neighboring silicon atoms, there will be one electron missing, forming a hole, as shown in Fig. 5.20. An electron with a small amount of energy in the valence band can move into the hole, and so the impurity is called an *acceptor*. From a band diagram perspective, the acceptor impurity energy level lies just above the valence bandedge, as shown in Fig. 5.21. For a silicon lattice, typical values of $E_a - E_v$ are on the order of 0.05 eV (e.g., if the dopant is boron, $E_a - E_v \simeq 0.45$ eV). Acceptor ionization levels can be determined from (5.45) using the effective mass of holes. The resulting material is known as a *p type* semiconductor.

TABLE 5.2 MEASURED VALUES OF DONOR IONIZATION ENERGIES FOR Si AND Ge (meV) ([6]).

	Ionization Energy (meV)		
Semiconductor	P	As	Sb
Ge	12.0	12.7	9.6
Si	45	49	39

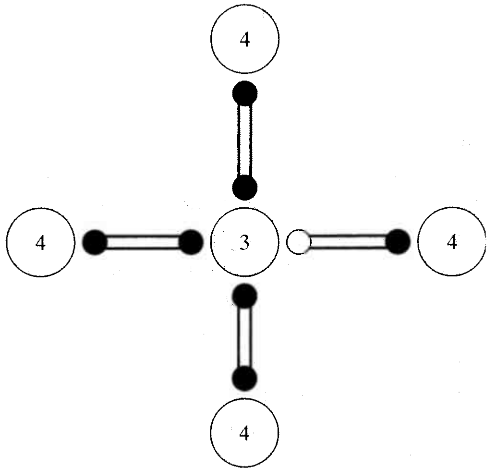


Figure 5.20 Part of a Group IV element lattice with the insertion of a Group III element. The double parallel lines indicate a covalent bond. One "extra" hole will be present, loosely bound to the Group III atom.

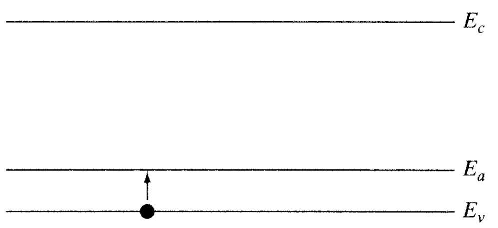


Figure 5.21 Real-space energy band diagram showing acceptor level.

5.4.2 Interacting Systems Model

There is another way to view bandstructure that is often helpful, especially in understanding how two systems interact when brought together. It turns out that if a quantum system has energy levels $E_1, E_2, E_3, \ldots$, then if two such identical systems (perhaps two atoms, or two quantum wells) are brought together, it can be shown that each energy level will split into two levels,

$$E_n \rightarrow E_n^+, E_n^-, \tag{5.46}$$

where $E_n^{\pm}$ is an energy value slightly above/below the energy value E_n of the isolated system. This is depicted in Figs. 5.22 and 5.23.

In Fig. 5.22, two identical systems, each having energy levels E_1 and E_2, are spaced sufficiently far apart so that they don't interact. Then, if the systems are brought near to

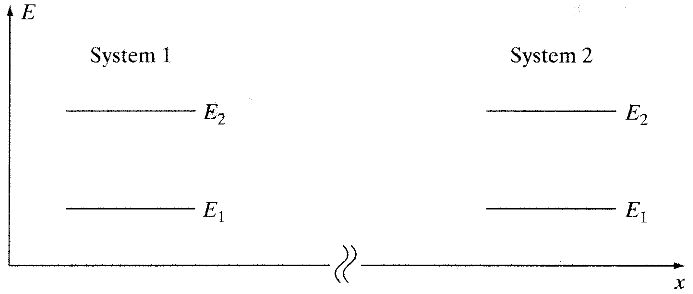


Figure 5.22 Two identical quantum systems, each having two energy levels, are spaced far apart. No interaction occurs.

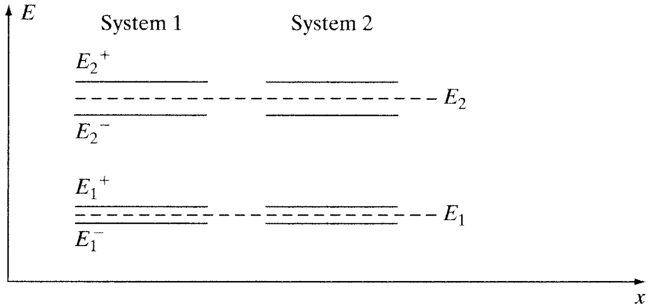


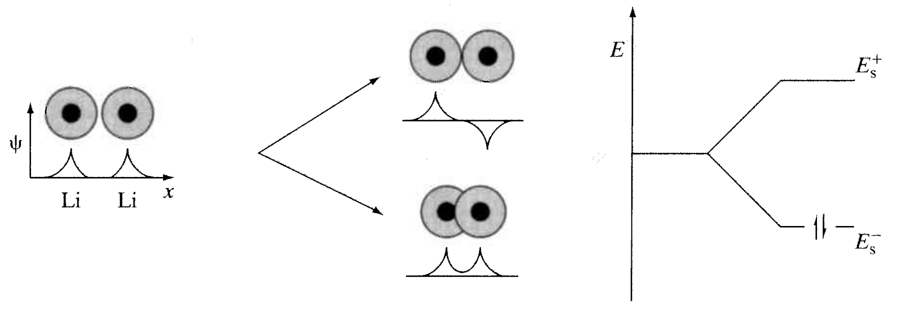
Figure 5.23 The two identical quantum systems from Fig. 5.22 are brought into close proximity, resulting in energy level splitting.

each other, as depicted in Fig. 5.23, the energy levels split as

$$E_1 \rightarrow E_1^+, E_1^-, \tag{5.47}$$

$$E_2 \rightarrow E_2^+, E_2^-. \tag{5.48}$$

The splitting is due to the overlap of each system's wavefunctions (really orbitals, as discussed in Section 3.5). For example, in the case of two atoms that come together to form a molecule, the atomic orbitals associated with each atom begin to overlap as the atoms are brought together. This can be seen by considering a simplified linear model of forming a lithium (Li) molecule. Lithium has the electronic configuration $1s^2 2s^1$, and in forming the molecule Li_2, the s shell atomic orbitals form *antibonding* and *bonding* molecular orbitals, as depicted in Fig. 5.24(a). In the ground configuration, the bonding molecular state is filled with the two $2s^1$ electrons (one from each atom), and the antibonding state is empty, as depicted in Fig. 5.24(b).



(a) (b)

Figure 5.24 Depiction of combining atomic orbitals of lithium to form molecular states of Li_2. (a) Two Li atoms are brought into close proximity. The atomic s shells interact, forming antibonding (top) and bonding (bottom) molecular orbitals. The $2s^1$ electron from each atom will go into the lowest energy state, as depicted in (b).

If N identical atoms are brought together, each energy level of an isolated atom, $E_1, E_2, E_3, \ldots$, will split into N levels,

$$E_1 \rightarrow \left\{ \begin{array}{c} E_{1,1} \\ E_{1,2} \\ \vdots \\ E_{1,N} \end{array} \right., \qquad E_2 \rightarrow \left\{ \begin{array}{c} E_{2,1} \\ E_{2,2} \\ \vdots \\ E_{2,N} \end{array} \right., \qquad E_3 \rightarrow \left\{ \begin{array}{c} E_{3,1} \\ E_{3,2} \\ \vdots \\ E_{3,N} \end{array} \right., \qquad \ldots, \tag{5.49}$$

where each group is centered on the corresponding isolated atomic level, E_n. As $N \rightarrow \infty$, the N discrete levels centered on each E_n merge into a quasi continuum of allowed energy levels, forming an energy band centered on E_n. If two adjacent levels of the original, widely spaced system, say, E_n and E_{n+1}, are well separated in energy, then the resulting energy bands centered on E_n and on E_{n+1} will not overlap, creating a bandgap. This is depicted in Fig. 5.25, where the s and p states of an atomic shell in an atom (Section 4.6.2) split into separate s and p bands in a solid. The top half of either resulting band (s or p in Fig. 5.25) is the antibonding half, and the bottom half is the bonding half.

If, for example, the atomic s shell in Fig. 5.25 is partially filled (i.e., an ns^1 shell), then the resulting s band will be partially filled (the bonding half), resulting in a conductor. If the shells in question are full (e.g., ns^2), then the resulting bands will be full, usually resulting in an insulator or semiconductor.

However, the situation is often much more complicated. For instance, assume that the s shell is full, and that the next p shell is empty. The resulting s band will be full, and the p band will be empty. However, these bands may overlap, allowing conduction. This

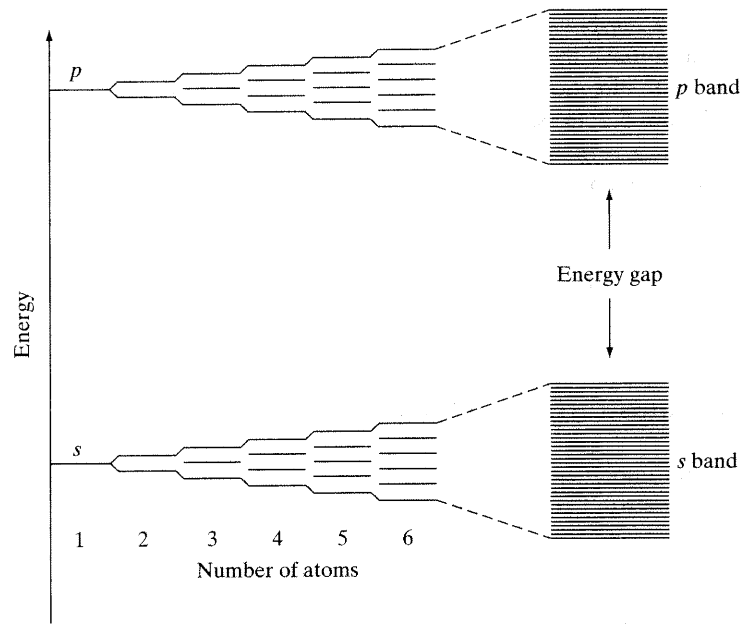


Figure 5.25 Depiction of s and p states in an atomic shell splitting into s and p bands as atoms are brought together to form a solid.

happens in, for example, Mg, which has the atomic configuration[†] $1s^2 2s^2 2p^6 3s^2 3p^0$ (i.e., full s shell, empty p shell) and yet is a metallic solid due to band overlap. Moreover, it is often energetically favorable for atomic states to become mixed (*hybridized*) in forming atomic bonds. This often occurs when the s and p shells are sufficiently close in energy such that the atomic orbitals can interact, resulting in sp molecular orbitals. This is extremely common, and occurs in most semiconductors.

The idea of combining atoms to form molecules and solids can be made quantitatively accurate by a method known as the linear combination of atomic orbitals (LCAO), although this is beyond the scope of the text. Nevertheless, it is an intuitively appealing model, since in the limit of large separation we expect that systems will not interact, but will begin to interact as they are brought closer together.

5.4.3 The Effect of an Electric Field on Energy Bands

In order to analyze electronic devices, we must be able to understand the effect of applying an electric field or a potential difference to a material. For simplicity, we will consider a one-dimensional example and a constant electric field $\mathbf{E} = \mathbf{a}_x \mathcal{E}_0$. From basic electrostatics, the work done (energy expended) in moving a charge q from 0 to x in the presence of the field is

$$W(x) = -\int_0^x \mathbf{F} \cdot d\mathbf{l} = -\int_0^x (\mathbf{a}_x q \mathcal{E}_0) \cdot \mathbf{a}_x dx = -q\mathcal{E}_0 x, \tag{5.50}$$

where $\mathbf{F} = q\mathbf{E}$ is the force on the charge q by the field[‡] $\mathbf{E}$. For an electron, $q = -e < 0$, and energy $-q_e \mathcal{E}_0 x = e\mathcal{E}_0 x$ is required to move the electron against the field. The energy required to move the electron to the point x increases linearly as x increases. In moving in the opposite direction, the electron gives up energy to the field.

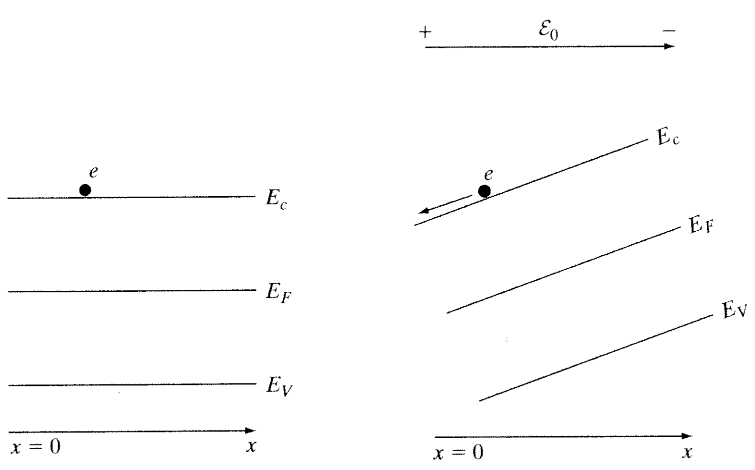
Figure 5.26(a) shows an electron in the conduction band of a material. To accommodate the idea of needing to linearly increase the electron's energy if it is to move in the positive x-direction, and decrease its energy in moving in the negative x-direction, we tilt the energy band diagram under the influence of the applied field, as shown in Fig. 5.26(b). We can think of the tilted bands as encouraging the electrons to "roll downhill," or requiring an energy input to climb uphill. From a voltage standpoint, positive voltage depresses (pushes down) the energy level.

5.4.4 Bandstructures of Some Semiconductors

Given the importance of semiconductors in electronics applications, we will very briefly examine the band structure of several types of semiconductors.

[†] $3p^0$ and higher shells are not usually listed, although it was included here to emphasize the point that the s shell is full and the next p shell is empty.

[‡] Since voltage is work per unit charge, $V(x) = W(x)/q$ (referenced to $V(0) = 0$); if we know the potential V rather than the electric field $\mathcal{E}$, energy is simply qV. Be aware that the symbol V can represent voltage or potential energy, although usually its meaning is clear from the context.



(a) (b)

Figure 5.26 (a) Real-space band diagram in the absence of an applied electric field. (b) Real-space band diagram in the presence of an applied field $\mathcal{E}_0$.

For most semiconductors, the top of the valence band occurs at $k = 0$. However, the bottom of the conduction band may occur at some other value of k, and we can divide semiconductors into two general categories.

Direct bandgap semiconductors are such that the bottom of the conduction band occurs at $k = 0$, as shown in Fig. 5.27, where E_c and E_v are the conduction and valence bandedges, respectively, and where the bandgap is $E_g = E_c - E_v$. For a direct bandgap semiconductor, near $k = 0$ the energy bands can be modeled by

$$\begin{aligned} E(k) &\simeq E_c + \frac{\hbar^2 k^2}{2m_e^*}, && \text{conduction band} \\ E(k) &\simeq E_v - \frac{\hbar^2 k^2}{2m_h^*}, && \text{valence band.} \end{aligned} \tag{5.51}$$

Materials such as GaAs and InP are direct bandgap semiconductors. (The bandstructure for GaAs is shown in Fig. 5.12 on p. 143.)

Indirect bandgap semiconductors are semiconductors for which the bottom of the conduction band does not occur at $k = 0$, but at some other wavenumber value, as shown in Fig. 5.28. Semiconductors such as Si, Ge, and AlAs are indirect bandgap semiconductors. (The bandstructure for Ge and Si is also shown in Fig. 5.12.) Actually, silicon has a complicated bandstructure, and the bottom of the conduction band occurs at six equivalent minima in k-space, resulting in six conduction band valleys.

Indirect semiconductors highlight the difference between crystal momentum and particle momentum, since at the bottom of the conduction band in an indirect semiconductor, the particle momentum is zero, although the crystal momentum $\hbar k$ is nonzero. Particle momentum notwithstanding, the electron dynamics are governed by the equation of motion

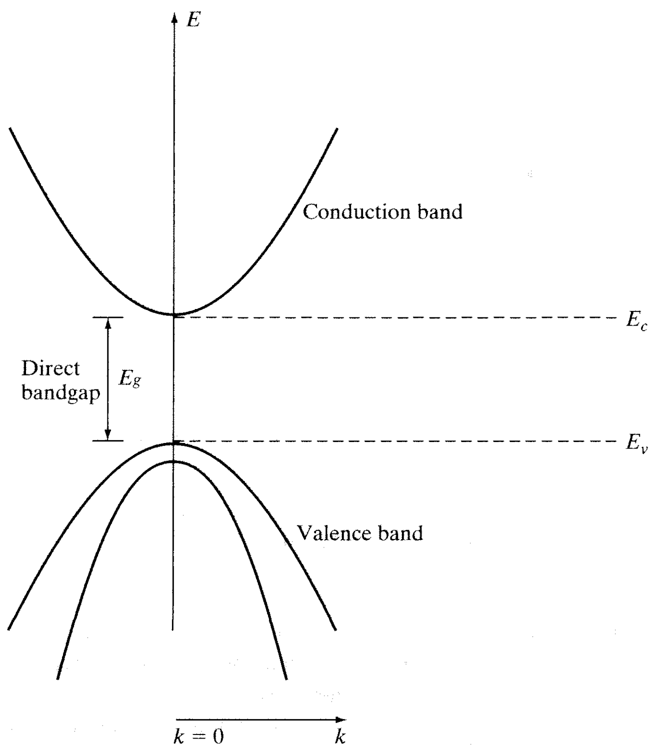


Figure 5.27 Energy bands (wavenumber space) for a typical direct bandgap material (light and heavy hole valance bands are shown).

involving the crystal momentum (5.34). The bandgap properties of several important semiconductors are given in Table 5.3.

Considering the conduction and valence bands, it should be noted that electron kinetic energy is given by $E - E_c > 0$, and hole kinetic energy by $E_v - E > 0$, as depicted in Fig. 5.29.

5.4.5 Electronic Band Transitions—Interaction of Electromagnetic Energy and Materials

So far, we have been considering primarily materials in thermal equilibrium, where electrons and holes have, on average, thermal energy corresponding to the temperature of the material. When an electron is thermally excited from a lower state (the valence band, or an impurity level) to the conduction band, the electron is available to contribute to conduction. In general, the electron will exist in the conduction band for some period of time before falling to a lower state, a process called *recombination* (as in an electron recombining with a hole). For a material in thermal equilibrium, the generation rate of carriers is equal to the recombination rate, so that the carrier concentration remains constant.

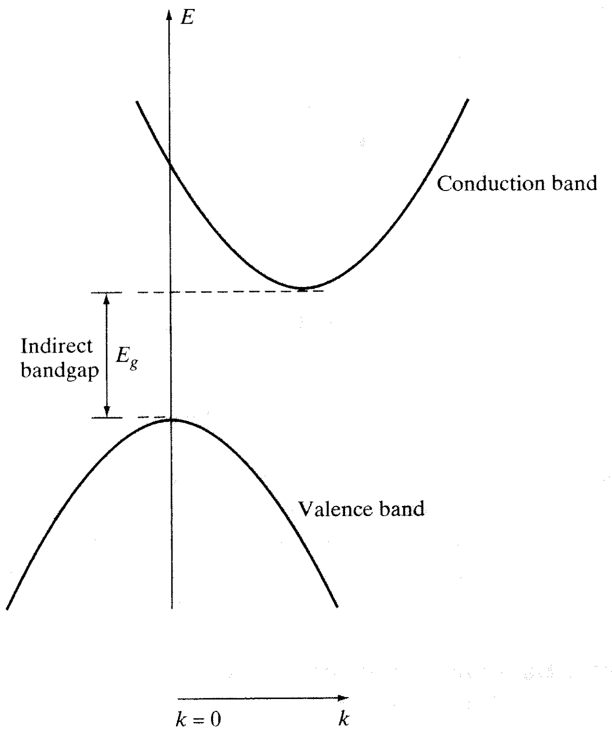


Figure 5.28 Energy bands (wavenumber space) for a typical indirect bandgap material.

TABLE 5.3 BANDGAP PROPERTIES OF SEVERAL IMPORTANT SEMICONDUCTORS. FOR GAP TYPE, I INDICATES AN INDIRECT BANDGAP SEMICONDUCTOR, AND D A DIRECT GAP SEMICONDUCTOR.

Crystal	Gap type	E_g (eV) @ 0 K	E_g (eV) @ 300 K
Si	I	1.17	1.11
Ge	I	0.74	0.66
GaAs	D	1.52	1.43
InP	D	1.42	1.27
AlAs	I	2.23	2.16
CdS	D	2.58	2.42
CdSe	D	1.84	1.74
CdTe	D	1.61	1.44

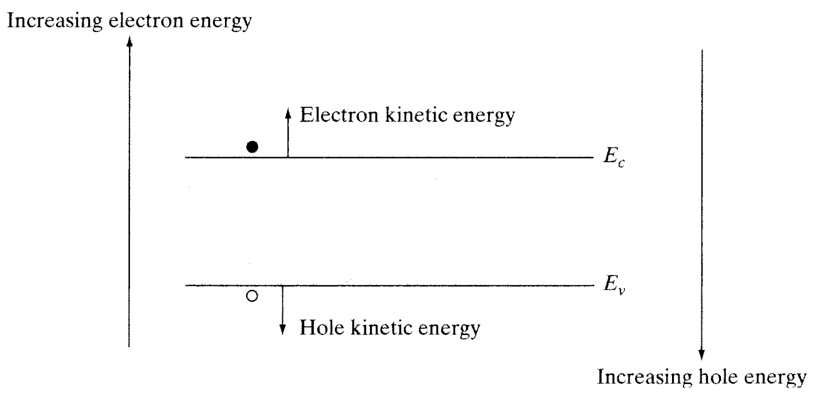
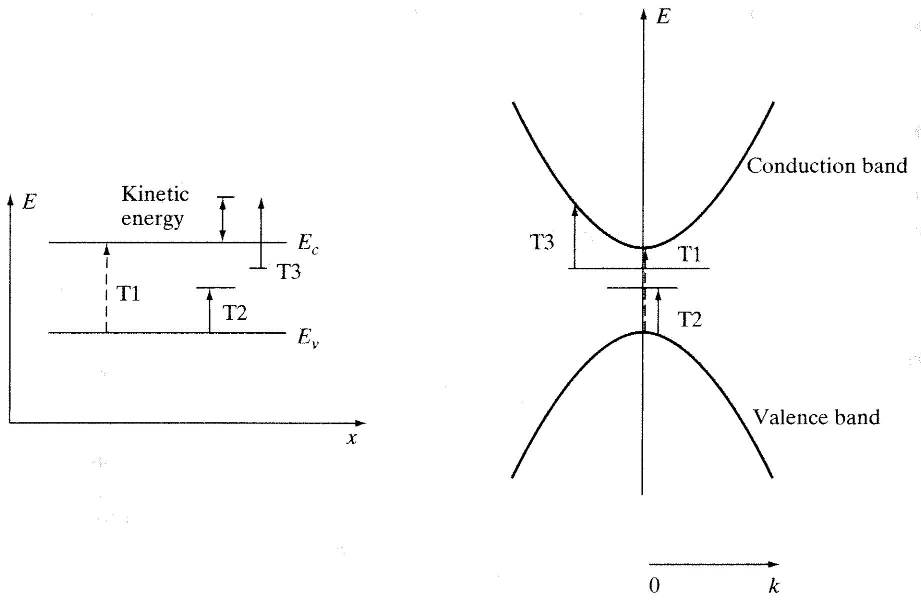


Figure 5.29 Depiction of electron and hole energy.

Nonequilibrium conditions refer to the presence of charge carriers by other means, typically by either direct electrical injection, or by electromagnetic processes (e.g., photons incident on the material). Since the interaction of electromagnetic energy and materials is obviously very important in electrical applications, in this section, a brief survey of several important interaction mechanisms will be provided.

Direct Bandgap Semiconductors. Figure 5.30 shows typical energy band diagrams for a direct bandgap material, where (a) depicts the energy bands in real space and (b) shows the energy bands in wavenumber space. Three possible transitions arising from the interaction of a photon and the material are depicted, denoted[†] as T1–T3. In all transitions, we must have conservation of energy and momentum.



(a) (b)

Figure 5.30 (a) Several possible band transitions as shown on the real-space band diagram. (b) The same transitions in wavenumber space.

T1 involves a transition from the valence bandedge (or below) to the conduction bandedge (or above). It is a *direct transition*, involving absorption of a photon. To elevate an electron from the valence bandedge to the conduction bandedge requires an energy input of $E_g = E_c - E_v$. Energy conservation requires that

$$\hbar\omega = E_g,$$

for a bandedge-to-bandedge transition (at $k = 0$), where $\hbar\omega$ is the energy of the incident photon. Photons having larger energies can cause larger transitions (i.e., perhaps from deeper in the valence band, or to deeper in the conduction band, or both). This can result in particles gaining kinetic energy (the excess energy measured from the bandedge), or energy going into lattice vibrations.

For the direct vertical transition T1 shown, conservation of momentum, which, via de Broglie's relation (2.15) corresponds to conservation of wavevector, states that

$$\Delta k = k_{pt},$$

where Δk is the change in the electron's wavevector, and k_{pt} is the wavevector of the photon. For optical photons, where $\lambda_{pt} \sim 400$–700 nm, the photon wavevector is in the range

$$k_{pt} = \frac{2\pi}{\lambda} \sim 1.5 \times 10^7 - 9.0 \times 10^6 \; m^{-1}.$$

However, electron wavevectors tend to be much larger. For example, for a crystal with lattice constant $a = 0.5$ nm, at the Brillouin zone boundary, the electron's wavevector is

$$k_e = \frac{\pi}{a} = 6.28 \times 10^9 \; m^{-1}.$$

So k_{pt} is rather small on the scale of the first Brillouin zone, and so, approximately, $\Delta k = 0$. Thus, we get the vertical transition shown as T1 in Fig. 5.30(b). This type of transition can be used to measure the bandgap of a direct bandgap semiconductor, by determining the smallest energy of photons that are absorbed by the semiconductor.

Transitions T2 and T3 represent possible transitions associated with impurity states. These impurities can be defects in the lattice, donor or accepter atoms, or other elements that may be present. This leads to the creation of *impurity states* in the bandgap, localized in positions that physically correspond to the location of the impurity. Therefore, the uncertainty in position of the impurity (say, Δx) is small, and, via the uncertainty principle, the uncertainty in Δk will be large. This is why the impurity level is depicted as being spread

---

[†] The same types of transitions can occur thermally, as previously discussed. However, considering that room temperature thermal energy is approximately 25 meV, only relatively small thermal transitions are likely.

over a relatively wide range of wavenumbers in Fig. 5.30(b), but over a small range of positions in Fig. 5.30(a).

Of course, although we have considered absorption of a photon[†] and elevation of an electron to a higher electronic state, the reverse process can also occur. That is, an electron can fall from an elevated state to a lower energy state by emitting a photon. This is the idea behind, for example, laser operation and material fluorescence. In fact, often an electron is raised to a point higher in the conduction band by absorption of a photon, and rapidly falls down to the bottom of the band by emitting quantized lattice vibrations (phonons), a process that is known as *nonradiative relaxation*. The electron then may fall from the bottom of the conduction bandedge to the valence band, by, for example, emitting a photon. Other types of transitions are also possible.

Indirect Bandgap Semiconductors. When one considers the interaction of energy and an indirect bandgap semiconductor, such a Si or Ge, the model changes somewhat. As depicted in Fig. 5.31, the transition from the top of the valence band to the bottom of the conduction band (i.e., the minimum energy transition) cannot be a vertical transition in wavenumber space. Because of the small wavevector associated with photons, photons enable essentially vertical transitions, providing energy but not enough momentum (wavenumber) for the necessary indirect transition. Some other interaction will be necessary for the transition to occur, and the answer is provided by phonons, which provide the necessary change in wavevector. Phonons have relatively large wavevectors, although their energy is typically low, and so phonons result in, essentially, horizontal transitions. Therefore, we have a three "particle" interaction: the photon, the phonon, and the electron. In a classical model, we would view the transition as arising from the interaction of the electromagnetic field, the electron, and the vibrating lattice. In the quantum model, we quantize the electromagnetic field as a photon, we quantize the vibrating lattice as a phonon, and the mutual interaction of these with the electron results in the transition, as depicted in Fig. 5.31. Conservation of energy and momentum results in

$$\Delta E = \hbar\left(\omega_{pt} \pm \omega_{pn}\right), \tag{5.52}$$

$$\Delta k = k_{pt} \pm k_{pn} \simeq \pm k_{pn}, \tag{5.53}$$

where $\hbar\omega_{pt}$ is the energy of an absorbed photon, and $\hbar\omega_{pn}$ is the energy of an absorbed (+) or emitted (−) phonon.

Note that the absorption of the photon alone puts the electron in the band gap, where, based on the previous theory, there are no electron states. A more sophisticated analysis would show that virtual states can exist within the bandgap. These states don't correspond to propagating waves, but decay exponentially with time, so that the electron must interact with a phonon very quickly in order to scatter into the conduction band. This makes the transition rate for indirect transitions much smaller than for direct transitions.

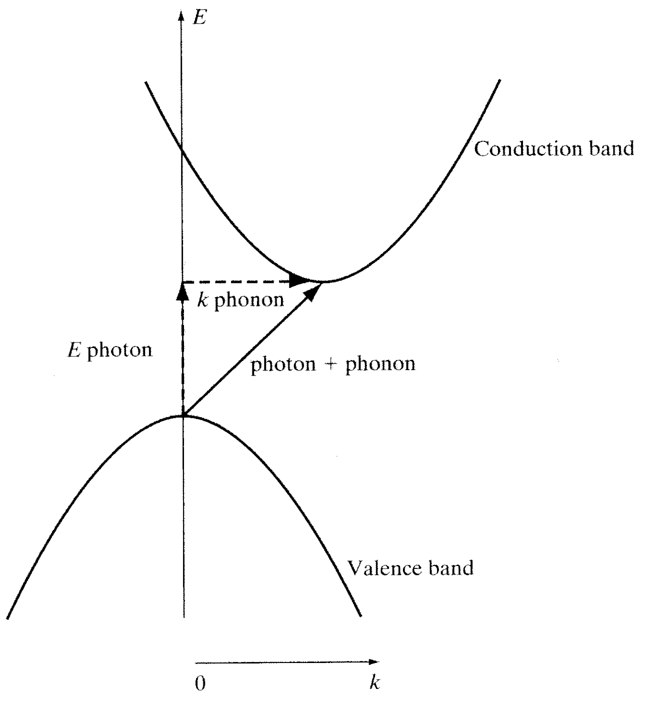


Figure 5.31 Transition in an indirect semiconductor involving a phonon and a photon. Because of the small wavevector associated with the photon, a phonon is required to provide the necessary change in wavevector.

Aside from bandedge-to-bandedge absorption as considered earlier, even in indirect semiconductors electrons can be elevated from the valence band directly to the conduction band at the same k value (i.e., by a vertical transition), although obviously this will occur at relatively high energies (higher then the bandgap energy). Although we won't discuss the transition rate in detail, there is a way of calculating how likely a transition is to occur. This is given by *Fermi's golden rule*, which states that the transition rate is proportional to a factor relating to the wavevectors of the two states in question, and to the joint density of states. (Density of states is discussed in Section 8.1.) In particular, over portions of the E–k plot where conduction and valence bands are essentially parallel, the same photon energy E can result in transitions for many different k values (thus leading to a large joint density of states), and the transition rate will be relatively large. For example, it can be seen from Fig. 5.12 on page 143 that silicon has a conduction and valence band that are approximately parallel, although the separation is on the order of 3–4 eV. Thus, one would expect absorption to greatly increase in the vicinity of photon energies of 3–4 eV, and this is indeed found to be the case. The direct absorption in this energy range is much larger than the indirect absorption near the bandgap energy.

[†] It is worthwhile to note that the idea of energy level transitions explains why materials are opaque at some frequencies and transparent at others. If the incident photon energy does not result in any energy transitions (and here we need to consider not just the interaction of photons and electrons, but also the interaction of photons and other quasi particles such as phonons), then the photon is not absorbed. The photon can therefore pass through the material, i.e., the material is transparent to that energy. This explains, for example, why a material that is opaque to visible light will usually be relatively transparent to lower frequency (lower-energy) electromagnetic waves, and thus, although we can't see through walls, we can receive cell phone calls in buildings.

Excitons. We have considered band transitions in semiconductors, and it was implicitly assumed that the process of absorption (of a photon, thermal energy, etc.) created a free electron and a free hole, each of which can contribute to conduction. There is another effect that is worth mentioning, primarily because of its importance in quantum-confined structures such as carbon nanotubes and quantum dots. The basic idea is that after an electron transition, it is possible for the electron and the created hole to be bound together by their mutual Coulomb attraction, forming a quasi particle known as an *exciton*. (See Chapter 9 for a discussion of excitons in quantum-confined structures.)

The two-particle electron-hole exciton can be modeled like the two-particle hydrogen atom considered in Section 4.6.1. However, unlike the hydrogen atom, which consists of one proton and one electron (having greatly different masses) in empty space, here the bound electron–hole pair moves through a material characterized by relative permittivity[†] ε_r, as depicted in Fig. 5.32.

Considering the formulas for energy and Bohr radius of the hydrogen atom, (4.115) and (4.117), and substituting $\varepsilon_r\varepsilon_0$ in place of ε_0 and the reduced mass m_r^*,

$$m_r^* = \frac{m_e^* m_h^*}{m_e^* + m_h^*}, \tag{5.54}$$

in place of m_e, then the binding energy and radius of the ground state exciton are given by

$$E = -\frac{m_r^* q_e^4}{8\varepsilon_r^2\varepsilon_0^2 h^2 n^2} = -\frac{m_r^*}{m_e\varepsilon_r^2}R_Y = -\frac{m_r^*}{m_e\varepsilon_r^2}13.6\text{ eV}, \tag{5.55}$$

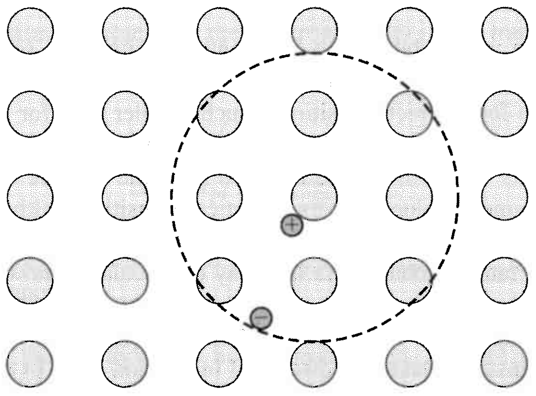


Figure 5.32 Depiction of a bound electron–hole pair known as an exciton. The exciton radius is much larger than the lattice constant.

[†] The macroscopic model of permittivity is appropriate here since the average separation of the electron and the hole is much greater than the lattice constant.

$$a_{ex} = \frac{4\pi\varepsilon_r\varepsilon_0\hbar^2}{m_r^* q_e^2} = \frac{\varepsilon_r m_e}{m_r^*}a_0 = \frac{\varepsilon_r m_e}{m_r^*}(0.53\text{ Å}), \tag{5.56}$$

where R_Y is the Rydberg energy and a_0 is the Bohr radius.

As an example, for GaAs ($\varepsilon_r = 13.3$), using an average of the heavy and light hole masses, $m_r^* = 0.0502m_e$, we find that

$$E = -3.86\text{ meV}, \tag{5.57}$$

$$a_{ex} = 265a_0 = 14\text{ nm}. \tag{5.58}$$

The binding energy of the pair, E, can be easily overcome by thermal effects (e.g., $k_BT \simeq 25$ meV at room temperature), thus breaking the exciton into free electrons and holes. Therefore, in bulk materials, exciton effects are usually only observed at very low temperatures,[†] and for relatively pure samples (since impurities such as dopants tend to screen the Coulomb interaction, much as occurs in conductors).

To demonstrate the concept of optical absorption by bandgap transitions and excitons, the absorption coefficient[‡] versus incident photon energy for GaAs is shown in Fig. 5.33 for various temperatures. At room temperature, one can observe the lack of absorption below

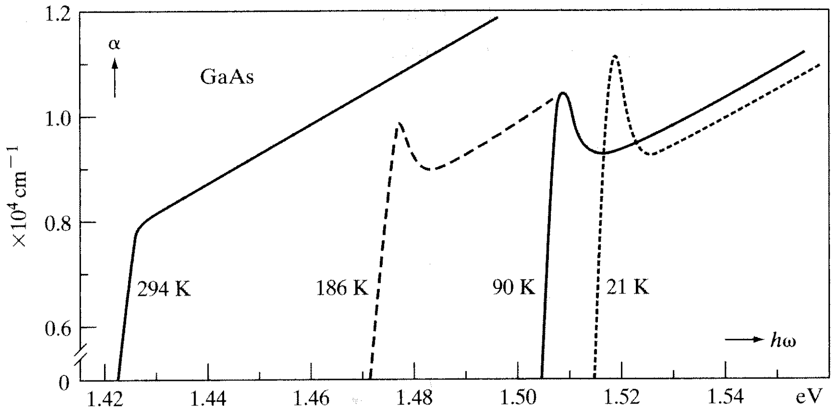


Figure 5.33 Absorption coefficient versus incident photon energy for GaAs at various temperatures. At room temperature, one can observe the lack of absorption below the bandgap. At low temperatures, the peaks at the onset of absorption are due to excitons. (Based on Figure 6.8 (page 55) from *Band Theory and Electronic Properties of Solids* by John Singleton, Oxford University Press, 2001. Reprinted with permission from John Singleton and Oxford University Press. Data from M.D. Sturge, *Phys. Rev.* **127**, 768 (1962).)

[†] Here we consider "free" excitons, known as *Wannier–Mott excitons*, which occur in semiconductors. Another type, known as *Frenkel excitons*, have a much smaller radius and can be stable at room temperature, although these are not of interest here.

[‡] The absorption coefficient is defined to be the fraction of power absorbed per unit length of a material.

the bandgap ($E_g = 1.43$ eV). At low temperatures, the peaks at the onset of absorption are due to the creation of excitons.

5.5 GRAPHENE AND CARBON NANOTUBES

Since the relatively recent discovery of *carbon nanotubes* (CNs), there has been an enormous amount of research into their fundamental properties, and great excitement concerning their possible applications. Electronic applications of CNs will be discussed later (e.g., in Sections 6.3.1, 7.3.1 and 10.3), and here we will concentrate on their band structure.

A single-wall carbon nanotube (SWNT) is, roughly speaking, a rolled-up sheet of graphene, which is a mono-atomic layer of graphite. CNs typically have radius values of a few nanometers, and lengths (so far) up to centimeters. Multiwalled carbon nanotubes (MWNTs) are also common, and other related structures exist, such as nanotube ropes (bundles of nanotubes), although here we will focus on SWNTs.

5.5.1 Graphene

At an atomic level, graphene has the periodic honeycomb structure shown in Fig. 5.34, where the small circles denote the location of carbon atoms and the lines represent carbon–carbon bonds. The depicted lattice basis vectors are $\mathbf{a}_1 = \left(\sqrt{3}\mathbf{a}_x + \mathbf{a}_y\right) a/2$ and $\mathbf{a}_2 = \left(\sqrt{3}\mathbf{a}_x - \mathbf{a}_y\right) a/2$, where $a = \sqrt{3}b$ and $b = 0.142$ nm is the interatomic distance between carbon atoms in graphene.

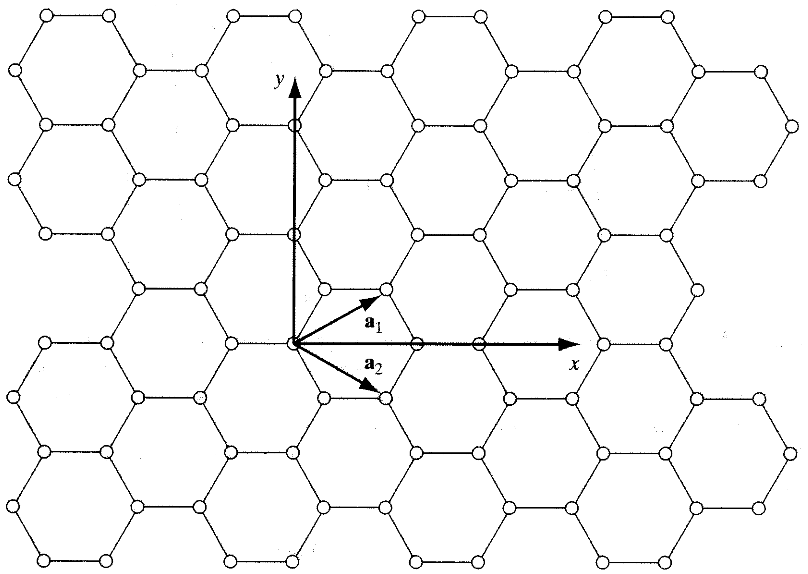
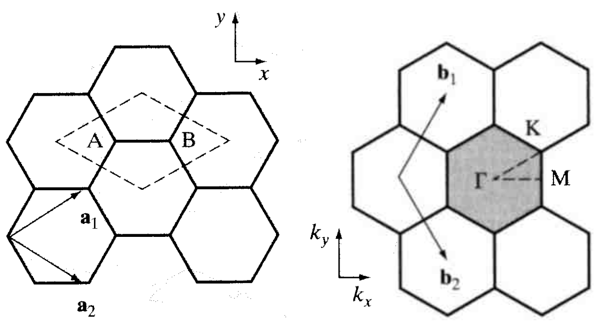


Figure 5.34 Graphene sheet (single layer of graphite). The small circles denote the location of carbon atoms. Lattice basis vectors are $\mathbf{a}_1$ and $\mathbf{a}_2$, as shown.



(a) (b)

Figure 5.35 (a) The unit cell (dotted rhombus) in graphene. (b) The first Brillouin zone in graphene (shaded region). Both the real-space and reciprocal space structure in graphene consists of hexagons. (Based on a figure in *Physical Properties of Carbon Nanotubes* by R. Saito, G. Dresselhaus, and M.S. Dresselhaus. Singapore: World Scientific Publishing Co. Pte. Ltd., 1998. Used by permission.)

Since graphene is a two-dimensional periodic material, it has an energy band structure similar to the three-dimensional crystalline solids discussed previously. The unit cell and first Brillouin zone are shown in Fig. 5.35, where the high-symmetry points in the Brillouin zone are

$$\Gamma = (0, 0), \quad \mathbf{K} = \left(\frac{2\pi}{\sqrt{3}a}, \frac{2\pi}{3a}\right), \quad \mathbf{M} = \left(\frac{2\pi}{\sqrt{3}a}, 0\right). \tag{5.59}$$

The Fermi surface is really a collection of points, the six points of the hexagonal Brillouin zone, where $E = E_F = 0$. (See problem 5.21.)

The most important bands arise from the so-called π–orbitals, and for these bands, the two-dimensional E–k relationship is well approximated by

$$E\left(k_x, k_y\right) = \frac{\mp\gamma_0 w\left(k_x, k_y\right)}{1 \pm s w\left(k_x, k_y\right)}, \tag{5.60}$$

where

$$w\left(k_x, k_y\right) = \sqrt{1 + 4\cos\left(\frac{\sqrt{3}k_x a}{2}\right)\cos\left(\frac{k_y a}{2}\right) + 4\cos^2\left(\frac{k_y a}{2}\right)}, \tag{5.61}$$

$s \simeq 0.129$, $\gamma_0 \simeq 3$ eV, and where the upper and lower signs correspond to the bonding/antibonding bands. (See Section 5.4.2.) In the important vicinity of the Fermi points, this can be approximated as

$$E\left(k_x, k_y\right) = \pm\gamma_0 w\left(k_x, k_y\right), \tag{5.62}$$

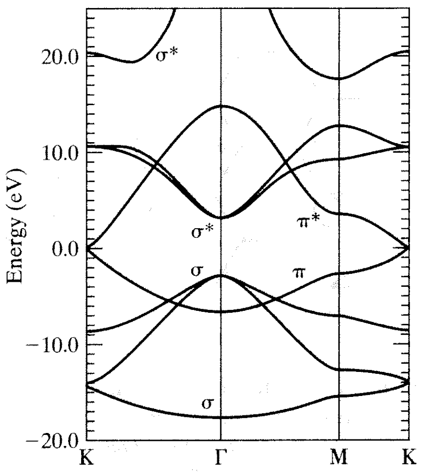


Figure 5.36 Calculated bandstructure of the π- and σ-bands in graphene. The Γ point is $\mathbf{k} = (0, 0)$, and the K and M points are shown in Fig. 5.35 (Based on a figure in *Physical Properties of Carbon Nanotubes* by R. Saito, G. Dresselhaus, and M.S. Dresselhaus. Singapore: World Scientific Publishing Co. Pte. Ltd., 1998.)

and from either (5.60) or (5.62), it is easy to plot the energy band behavior of the π electrons. (See problem 5.21.) The π orbitals arise from electrons that are weakly bound to the carbon atoms. It is these electrons that are of most importance in electronic applications, since a small amount of energy can free them for conduction. Electrons more strongly held to the carbon atoms are known as σ electrons, and these play an important role at higher energies. The band structure of both π- and σ-electrons in graphene is shown in Fig. 5.36.

It is particularly important to note the nearly linear dispersion of the π bands near the K point. Hence, rather then the usual parabolic dispersion for an electron in free space or in a material, $E = \hbar^2 k^2/2m^*$, dispersion for electrons near the K point in graphene is

$$E = v_F \hbar k, \tag{5.63}$$

where $v_F \simeq 9.71 \times 10^5$ m/s is the Fermi velocity of electrons (to be discussed later). For photons, from (2.25),

$$E = c\hbar k, \tag{5.64}$$

where c is the speed of light, and so near the π-band crossing point electrons act more like photons than particles with mass. In this case the electrons are called *Dirac fermions*, and their properties are of interest in recently-developed graphene devices.

5.5.2 Carbon Nanotubes

A carbon nanotube is formed by wrapping the graphene sheet into a cylinder,[†] as depicted in Fig. 5.37, by connecting points O and A and B and B'. The circumference of the tube is related to the length of the *chiral vector*, $\mathbf{C}_h = n\mathbf{a}_1 + m\mathbf{a}_2$, where n, m are integers.

[†] Carbon nanotubes form naturally in, for example, the arc discharge of carbon electrodes, and are not made by literally rolling graphene sheets into cylinders.

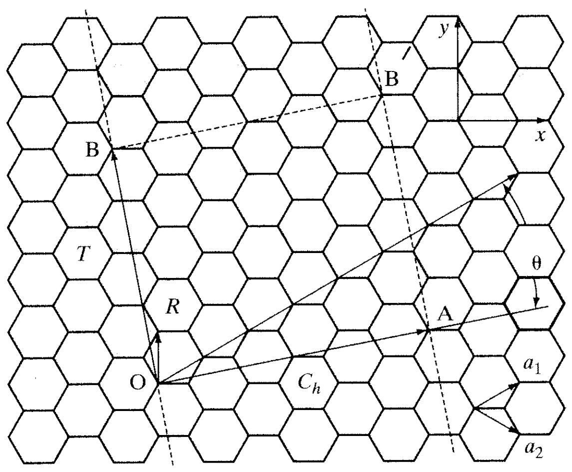


Figure 5.37 Depiction of forming a carbon nanotube by rolling a graphene sheet into a tube. (Based on a figure in *Physical Properties of Carbon Nanotubes* by R. Saito, G. Dresselhaus, and M.S. Dresselhaus. Singapore: World Scientific Publishing Co. Pte. Ltd., 1998. Used by permission.)

It is obvious that the CN cylinder can be formed by wrapping the graphene sheet along any preferred axis. If the cylinder axis is the x axis in Fig. 5.34, the resulting tube is called a *zigzag* CN because at an open end of the tube the carbon–carbon bonds would form a zigzag pattern. If the cylinder axis is the y axis in Fig. 5.34, the resulting tube is called an *armchair* CN, and if the cylinder axis is neither the x nor the y axis as shown, the resulting nanotube is called a *chiral* CN. Thus, carbon nanotubes can be characterized by the dual index (n, m), where $(n, 0)$ for zigzag CNs, (n, n) for armchair CNs, and (n, m), $0 < m \neq n$, for chiral nanotubes, as depicted in Fig. 5.38.

The values of n and m denote, respectively, the number of unit vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ required to make the tube. The resulting cross-sectional radius of a carbon nanotube is given by[†]

$$r = \frac{|\mathbf{C}_h|}{2\pi} = \frac{\sqrt{3}}{2\pi} b\sqrt{n^2 + nm + m^2}. \tag{5.65}$$

An armchair, zigzag, and chiral CN are shown in Fig. 5.39. Also, shown in the figure is a multi-wall CN.

Since graphene is a two-dimensional periodic structure, upon forming an infinite tube, we have a periodic structure in the axial direction and a finite structure in the transverse direction. It can be shown that the lattice constants are $a_{ac} = a = \sqrt{3}b$ for armchair tubes and $a_{zz} = \sqrt{3}a = 3b$ for zigzag tubes, and, therefore, the edge of the first Brillouin zone occurs at $k_y = \pi/a_{ac}$ for armchair tubes, and $k_x = \pi/a_{zz}$ for zigzag tubes. In the transverse direction, the wavenumber becomes quantized by the finite circumference of the tube, and

---

[†] The interatomic distance $b = 0.142$ nm in graphene becomes slightly larger in a carbon nanotube, and the value $b = 0.144$ nm is often used.

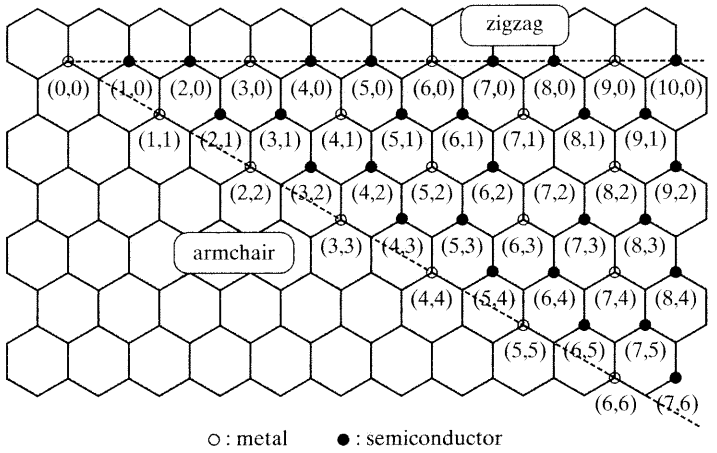


Figure 5.38 Forming armchair, zigzag, and chiral carbon nanotubes, denoted by the index (n, m). Metal and semiconducting tubes are denoted by hollow and solid circles, respectively. (Based on a figure in *Physical Properties of Carbon Nanotubes* by R. Saito, G. Dresselhaus, and M.S. Dresselhaus. Singapore: World Scientific Publishing Co. Pte. Ltd., 1998. Used by permission.)

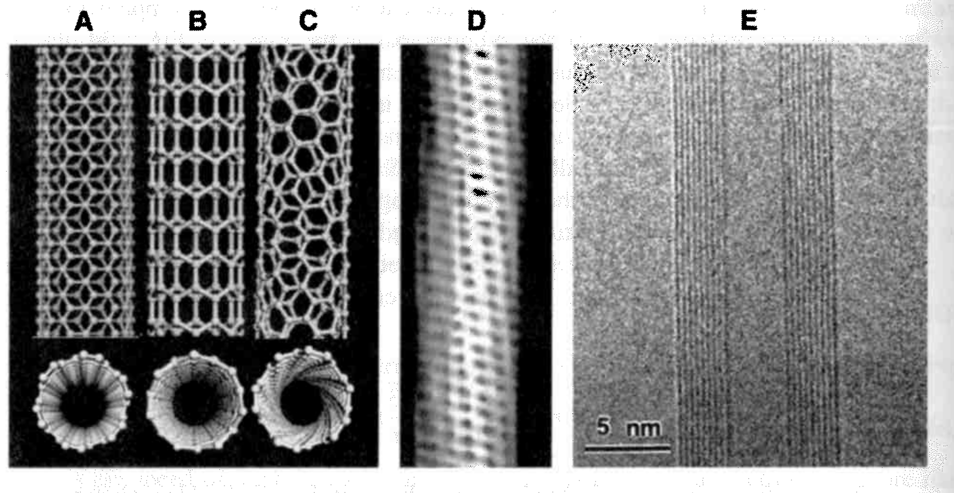


Figure 5.39 (A) armchair CN, (B) zigzag CN, (C) chiral CN, (D) TEM image of a 1.3 nm diameter chiral CN, (E) TEM image of a multi-wall CN, consisting of nine concentric single-wall CNs. (Courtesy Shenzhen Nano-Technologies Port Co., Ltd.)

is given by

$$k_{\perp} = k_{x,q} = \frac{2\pi q}{n3b}, \qquad q = 1, 2, \ldots, 2n \tag{5.66}$$

for the armchair tube $(m = n)$, (see problem 5.23) and

$$k_{\perp} = k_{y,q} = \frac{2\pi q}{n\sqrt{3}b}, \qquad q = 1, 2, \ldots, 2n \tag{5.67}$$

for the zigzag tube. The $E - k$ relations for the π electrons in a nanotube result from plugging (5.66) or (5.67) into (5.62), such that for armchair tubes we have[†]

$$E_{ac}\left(k_y\right) = \pm\gamma_0\sqrt{1 + 4\cos\left(\frac{\pi q}{n}\right)\cos\left(\frac{k_y a}{2}\right) + 4\cos^2\left(\frac{k_y a}{2}\right)}, \tag{5.68}$$

$-\pi < k_y a < \pi$, $q = 1, 2, \ldots, 2n$, and for zigzag tubes,

$$E_{zz}\left(k_x\right) = \pm\gamma_0\sqrt{1 + 4\cos\left(\frac{\sqrt{3}k_x a}{2}\right)\cos\left(\frac{\pi q}{n}\right) + 4\cos^2\left(\frac{\pi q}{n}\right)}, \tag{5.69}$$

$-\pi < k_x\sqrt{3}a < \pi$, $q = 1, 2, \ldots, 2n$. These are approximate relations, since, among other things, the effect of tube curvature on atomic bonds is ignored, although (5.68) and (5.69) turn out to be generally quite accurate. The dispersion behavior is considered in problem 5.24, although in general it is found that carbon nanotubes can be either metallic or semiconducting, depending on their geometry (i.e., on n, m).[‡] Armchair CNs are always metallic (they exhibit no energy bandgap), as are zigzag CNs with $n = 3p$, where p is an integer. Chiral tubes can be either metallic or semiconducting depending on the (n, m) values, and a general rule is that tubes where $(2n + m)/3$ is an integer are metallic.[§] Semiconducting tubes have energy bandgaps given by

$$E_g \simeq \frac{\gamma_0 a}{2\sqrt{3}r} \simeq \frac{0.383}{d_t}\ \text{eV}, \tag{5.70}$$

where d_t is the tube radius in nm.

Carbon nanotubes are held together by the carbon–carbon bonds, and thus they exhibit extraordinary strength. In fact, CNs are many times stronger than steel. They also have very high thermal conductivity and stiffness. Their rather extraordinary properties are in part due to the fact that they can be grown to be nearly defect free. (However, see the discussion in Section 10.3.) For example, metals tend to exhibit failure well below their theoretical limits,

---

[†] For graphene, $\gamma_0 \simeq 3$ eV. For carbon nanotubes, the corresponding value changes due to curvature effects, and is usually taken to be in the range 2.5–3.0 eV.

[‡] Graphene itself is a *semimetal* (also called a zero bandgap semiconductor), exhibiting properties between a metal and a semiconductor. This is why, merely depending on how the cylinder is wrapped, the resulting tube can be either metallic or semiconducting. Semiconducting CNs have bandgaps ranging from a few meV up to, on the order of, an eV, and metallic CNs tend to have very high conductivities.

[§] The described characterization of tubes as being metallic or semiconducting is based on the simple idea of rolling up a graphene sheet without, essentially, changing its atomic structure. This is valid for relatively large-radius tubes (above perhaps a nanometer), although for small-radius tubes the large radius of curvature hybridizes the π– and σ–orbitals, and deviations to the presented theory are encountered. For example, whereas a (5, 0) tube should be semiconducting based on the simple analysis, it is actually found to be metallic. Moreover, some zigzag tubes that should be metallic by the simple mode are actually semiconducting. See, e.g., [16].

Furthermore, the analysis assumes infinitely long tubes. Finite length tubes having lengths less then approximately 10 nm tend to act more like quantum dots (zero-dimensional structures) rather than quantum wires (one-dimensional structures).

due to defects. High quality (nearly molecularly perfect) CNs can be formed that perform near to their theoretical limit, which results in extremely attractive mechanical and electrical properties. Carbon nanotubes are the stiffest known fibers, and exhibit the highest tensile strength of any known material. The Young's modulus of CNs have been measured to be more than an order of magnitude larger than that for steel. Furthermore, carbon nanotubes can carry very high current densities (much higher than typical metals), at 10^9 A/cm^2 or more, without melting. For comparison, copper is generally limited to 10^6 A/cm^2 due to heating and electromigration effects.

For the semiconducting tubes, electronic properties can be controlled by doping, as in a conventional semiconductor. Doping can be introduced chemically by exposing the CN to certain elements, or by inserting molecules inside the tube. One inherent difference between semiconducting CNs and ordinary semiconducting materials is that at room temperature, the undoped CN tends to be a p-type material. It was first thought that this behavior was due to chemical contamination, such as adsorbed oxygen. However, it is now thought that the origin of the p-type behavior is due to what is known as *self-doping*, caused by the curvature of the tube at the nanoscale and the commensurate effect on atomic bonding.

Carbon nanotubes can currently be fabricated using a variety of techniques including carbon arc discharge, laser evaporation, and chemical vapor deposition. Often, however, a mixture of semiconducting and metallic tubes are produced, which must be separated and isolated for use. Making good electrical contact between carbon nanotubes and electrodes can be problematic, as is positioning of the tubes in device fabrication.

5.6 MAIN POINTS

This chapter presents some important concepts from solid state physics, and, in particular, the formation of energy bands in periodic structures. After studying this chapter you should know

- the principles of various crystal structures;
- the effect of a periodic potential on electron properties, and the Kronig–Penney model;
- the band theory of solids;
- the concept of effective mass, and what effective mass accounts for, including the use of effective mass in Schrödinger's equation;
- the effect of an electric field or of an applied electrical potential on energy band structure;
- energy band models for typical semiconductors, including the concept of direct and indirect bandgaps;
- the interaction of electromagnetic energy with an energy band system, including the role of vibrational modes (phonons);
- the basic π band structure of graphene;
- the band structure of π electrons in carbon nanotubes.

5.7 PROBLEMS

1. To gain an appreciation of the important role of surface effects at the nanoscale, consider building up a material out of bcc unit cells. (See Section 5.1.) For one bcc cube, there would be nine atoms, eight on the outside and one interior, as depicted on page 134. If we constrain ourselves to only consider cubes of material, the next largest cube would consist of eight bcc unit cells, and so on. If one side of the bcc unit cell is 0.5 nm, how long should the material's side be in order for there to be more interior atoms than surface atoms?
2. Consider the Kronig–Penney model of a material with $a_1 = a_2 = 5$ Å and $V_0 = 0.5$ eV. Determine numerically the starting and ending energies of the first allowed band.
3. Use the equation of motion (5.34) to show that the period of Bloch oscillation for a one-dimensional crystal having lattice period a is

$$\tau = \frac{h}{e\mathcal{E}a}, \tag{5.71}$$

where $\mathcal{E}$ is the magnitude of the applied electric field.

4. Determine the probability current density (A/m) from (3.187) for the Bloch wavefunction

$$\psi(x) = u(x)\, e^{ikx} e^{-i\omega t}, \tag{5.72}$$

where u is a time-independent periodic function having the period of the lattice,

$$u(x) = u(x + a). \tag{5.73}$$

5. If an energy–wavevector relationship for a particle of mass m has the form

$$E = \frac{\hbar^2}{3m} k^2, \tag{5.74}$$

determine the effective mass. (Use (5.29).)

6. If the energy–wavenumber relationship for an electron in some material is

$$E = \frac{\hbar^2}{2m} \cos(k), \tag{5.75}$$

determine the effective mass and the group velocity. (Use (5.29).) Describe the motion (velocity, direction, etc.) of an electron when a d.c. (constant) electric field is applied to the material, such that the electric field vector points right to left (e.g., an electron in free space would then accelerate towards the right). In particular, describe the motion as k varies from 0 to 2π. Assume that the electron does not scatter from anything.

7. If the energy–wavenumber relationship for an electron in some material is

$$E = E_0 + 2A\cos(ka), \tag{5.76}$$

determine the electron's position as a function of time. Ignore scattering.

8. Consider an electron in a perfectly periodic lattice, wherein the energy–wavenumber relationship in the first Brillouin zone is

$$E = \frac{\hbar^2 k^2}{5m_e}, \tag{5.77}$$

where m_e is the mass of an electron in free space. Write down the time-independent effective mass Schrödinger's equation for one electron in the first Brillouin zone, ignoring all interactions except between the electron and the lattice. Define all terms in Schrödinger's equation.

9. Assume that a constant electric field of strength $\mathcal{E} = -1$ kV/m is applied to a material at $t = 0$, and that no scattering occurs.

(a) Solve the equation of motion (5.34) to determine the wavevector value at $t = 1, 3, 7,$ and 10 ns.

(b) Assuming that the period of the lattice is $a = 0.5$ nm, determine which Brillouin zone the wavevector is in at each time. If the wavevector lies outside the first Brillouin zone, map it into an equivalent place in the first zone.

10. Using the hydrogen model for ionization energy, determine the donor ionization energy for GaAs ($m_e^* = 0.067 m_e$, $\varepsilon_r = 13.1$).

11. Determine the maximum kinetic energy that can be observed for emitted electrons when photons having $\lambda = 232$ nm are incident on a metal surface with work function 5 eV.

12. Photons are incident on silver, which has a work function $e\phi = 4.8$ eV. The emitted electrons have a maximum velocity of 9×10^5 m/s. What is the wavelength of the incident light?

13. In the band theory of solids, there are an infinite number of bands. If, at $T = 0$ K, the uppermost band to contain electrons is partially filled, and the gap between that band and the next lowest band is 0.8 eV, is the material a metal, an insulator, or a semiconductor?

14. In the band theory of solids, if, at $T = 0$ K, the uppermost band to have electrons is completely filled, and the gap between that band and the next lowest band is 8 eV, is the material a metal, an insulator, or a semiconductor? What if the gap is 0.8 eV?

15. Describe in what sense an insulator with a finite band gap cannot be a perfect insulator.

16. Draw relatively complete energy band diagrams (in both real space and momentum space) for a p-type indirect bandgap semiconductor.

17. For an intrinsic direct bandgap semiconductor having $E_g = 1.72$ eV, determine the required wavelength of a photon that could elevate an electron from the top of the valence band to the bottom of the conduction band. Draw the resulting transition on both types of energy band diagrams (i.e., energy–position and energy–wavenumber diagrams).

18. Determine the required phonon energy and wavenumber to elevate an electron from the top of the valence band to the bottom of the conduction band in an indirect bandgap semiconductor. Assume that $E_g = 1.12$ eV, the photon's energy is $E_{pt} =$ 0.92 eV, and that the top of the valence band occurs at $k = 0$, whereas the bottom of the conduction band occurs at $k = k_a$.

19. Calculate the wavelength and energy of the following transitions of an electron in a hydrogen atom. Assuming that energy is released as a photon, using Table 1, on page 4 classify the emitted light (e.g., x-ray, infrared (IR), etc.).

(a) $n = 2 \rightarrow n = 1$

(b) $n = 5 \rightarrow n = 4$

(c) $n = 10 \rightarrow n = 9$

(d) $n = 8 \rightarrow n = 2$

(e) $n = 12 \rightarrow n = 1$

(f) $n = \infty \rightarrow n = 1$

20. Excitons were introduced in Section 5.4.5 to account for the fact that sometimes when an electron is elevated from the valence band to the conduction band, the resulting electron and hole can be bound together by their mutual Coulomb attraction. Excitonic energy levels are located just below the band gap, since the usual energy to create a free electron and hole, E_g, is lessened by the binding energy of the exciton. Thus, transitions can occur at

$$E = E_g - \frac{m_r^*}{m_e \varepsilon_r^2} 13.6 \text{ eV}, \tag{5.78}$$

where E_g is in electron volts.[†]

(a) For GaAs, determine the required photon energy to create an exciton. For m_r^*, use the average of the heavy and light hole masses.

(b) The application of a d.c. electric field tends to separate the electron and the hole. Using Coulomb's law, show that the magnitude of the electric field between the electron and the hole is

$$|\mathcal{E}| = \left(\frac{m_r^*}{m_e}\right)^2 \frac{2}{\varepsilon_r^3 \, |q_e|} \frac{R_Y}{a_0}. \tag{5.79}$$

(c) For GaAs, determine $|\mathcal{E}|$ from (5.79). Determine the magnitude of an electric field that would break apart the exciton.

21. The E–k relationship for graphene is given by (5.62). The Fermi energy for graphene is $E_F = 0$, and the first Brillouin zone forms a hexagon (as shown in Fig. 5.35), the six corners of which correspond to $E = E_F = 0$. The six corners of the first Brillouin zone are located at

$$k_x = \pm \frac{2\pi}{\sqrt{3}a}, \quad k_y = \pm \frac{2\pi}{3a}, \tag{5.80}$$

and

$$k_x = 0, \quad k_y = \pm \frac{4\pi}{3a}. \tag{5.81}$$

[†] The quantity 13.6 should really be replaced by $13.6/n^2$, where n is the energy level of the exciton. Here we consider the lowest level exciton ($n = 1$), which is dominant.

(a) Verify that at these points, $E = E_F = 0$.

(b) At the six corners of the first Brillouin zone, $|\mathbf{k}| = 4\pi/3a$. Make a two-dimensional plot of the E–k relationship for k_x, k_y extending a bit past $|\mathbf{k}|$. Verify that the bonding and antibonding bands touch at the six points of the first Brillouin zone hexagon, showing that graphene is a semimetal (sometimes called a zero bandgap semiconductor). Also make a one-dimensional plot of $E\left(0, k_y\right)$ for $-|\mathbf{k}| \leq k_y \leq |\mathbf{k}|$, showing that the bands touch at $E = 0$ at $k_y = \pm 4\pi/3a$.

22. What is the radius of a $(19, 0)$ carbon nanotube? Repeat for a $(10, 10)$ nanotube. Consider an $(n, 0)$ zigzag carbon nanotube that has radius 0.3523 nm. What is the value of the index n?

23. Since carbon nanotubes are only periodic along their axis, the transverse wavenumber becomes quantized by the finite circumference of the tube. Derive (5.66) and (5.67) by enforcing the condition that an integer number q of transverse wavelengths must fit around the tube ($k_\perp = 2\pi/\lambda_\perp$).

24. Using (5.68) and (5.69), plot the dispersion curves for the first eight bonding and antibonding bands in a $(5, 5)$, $(9, 0)$, and $(10, 0)$ carbon nanotube. Let the axial wavenumber vary from $k = 0$ to $k = \pi/a_{ac}$ for the armchair tube, and from $k = 0$ to $k = \pi/a_{zz}$ for the zigzag tube. Comment on whether each tube is metallic or semiconducting, and identify the band (i.e., the q value) that is most important. If the tube is semiconducting, determine the approximate band gap.

Part

II

SINGLE-ELECTRON AND FEW-ELECTRON PHENOMENA AND DEVICES

In previous chapters, Schrödinger's equation and the principles of quantum physics were developed, with an emphasis on single particles (primarily electrons) and collections of noninteracting particles in different spatial regions. The remainder of the text is divided into two parts, and presents some basic nanoelectronic applications of these principles. In the next part, we will be concerned with physical phenomena associated with single electrons, or small numbers of electrons (perhaps, say, 10^0–10^5 electrons). The main emphasis is on electrons confined to nanoscopic spaces, such as quantum dots, and devices constructed from quantum dots and "charge islands." Nanoelectronics principles are developed for the so-called "single-electron" devices, including the single-electron transistor, after the important concept of Coulomb blockade has been discussed. Although most single-electron devices are at an early stage of development, especially in the area of manufacturability, they offer the potential benefits of ultralarge scale integration, with device dimensions on the order of nanometers. They also may exhibit very low power dissipation, and high speed. All of these positive attributes arise from the need to move only single electrons, or small groups of electrons, through devices.

The use of the term "single-electron" device merits some discussion. In conventional microelectronics, currents are typically on the order of 1 μA to 1 mA, corresponding to the movement of 6.25×10^6–6.25×10^9 electrons per microsecond. This occurs through a device perhaps 100 nm in length. Even considering devices at the upper limit of optical