

The $\cos ka$ is only defined within the limits between $+1$ and -1 , indicating that, since α is related to E , the solution for energies should also be between those limits. This essentially implies that the electrons, moving in a periodically varying potential field, may possess energies within certain energy bands only, i.e., within shaded regions in Fig. 3.8. These allowed energy bands are separated by ranges of αa corresponding to $\cos ka$ being either greater than $+1$ or less than -1 , i.e., corresponding to forbidden energy ranges. Thus, to summarize the main observations, (i) there are allowed and forbidden bands of energy for electrons moving through a periodic potential, (ii) the size of these bands varies as a function of P , i.e., $V_0 b$, and (iii) with increasing αa , i.e., with increasing E , the forbidden bands become narrower. Also, for larger P , the curve is steeper, which results in narrower allowed bands and wider forbidden bands. It is important to note that at the boundary of an allowed band $\cos ka = \pm 1$, and hence $k = n\pi/a$, indicating the discontinuities in energy occurring at these values of k . These conditions also correspond to the Bragg reflection rule, suggesting that the electron states with $k = n\pi/a$ can be described as standing waves, i.e., these electrons cannot propagate through the lattice, indicating the presence of the energy gap for that given k .

Two extreme cases for the energy variation as a function of wave vector are (i) the case of free electrons (for vanishing small product of $V_0 b$) and (ii) the case of bound electrons in a potential well (for $P \rightarrow \infty$). In the former case ($P \rightarrow 0$), $\cos \alpha a = \cos ka$, and hence, using Eq. (3.3.30)

$$E = \frac{\hbar^2 k^2}{2m_e} \quad (3.3.32)$$

i.e., the case of free electrons. In the case of $P \rightarrow \infty$, $\sin \alpha a = 0$, which is possible for $\alpha a = n\pi$, or $\alpha^2 = n^2 \pi^2 / a^2$ ($n = 1, 2, 3, \dots$). Using Eq. (3.3.30) ($\alpha^2 = 2m_e E / \hbar^2$), we can write

$$E_n = \frac{\hbar^2 \pi^2}{2m_e a^2} n^2 \quad (n = 1, 2, 3, \dots) \quad (3.3.33)$$

i.e., the case of electrons bound between two infinitely high potential barriers (for sufficiently thick barrier b , so that electrons with energies less than the barrier height V_0 cannot tunnel from one atomic site to the next). The analysis of these two limits indicates that by varying P from 0 to ∞ , one can obtain various cases from free electrons to the completely bound electrons. For intermediate values of P , energy bands (i.e., ranges of allowed energies) are formed and these are separated by forbidden gaps (or energy gaps).

3.4. ENERGY BANDS IN CRYSTALS

In general, there are two ways to describe the physical properties of solids. One requires no need of invoking the periodic potential; instead, the description is based on the chemical bonds in the material. The other method involves the description

of the properties of electrons in the long-range periodic potential. The existence of the energy gaps, and of other related properties, can be considered on the basis of the energy required to remove an electron from a chemical bond in the material, and to allow it to freely move through the material (under the applied field). A more rigorous description of the energy bands in crystals requires the derivation of the $E(k)$ relationship, which facilitates the elucidation of important characteristics of the electronic properties of semiconductors. For free electrons, this relationship is

$$E = \frac{\hbar^2 k^2}{2m_e} \quad (3.4.1)$$

where

$$k = (2m_e E / \hbar^2)^{1/2} \quad (3.4.2)$$

which for one-dimensional case gives

$$k_x = (2m_e / \hbar^2)^{1/2} E^{1/2} \quad (3.4.3)$$

which is a parabolic function (see Fig. 3.1).

From the Kronig–Penney model

$$\frac{P \sin \alpha a}{\alpha a} + \cos \alpha a = \cos ka \quad (3.4.4)$$

and for free electrons (i.e., $P = 0$),

$$\cos \alpha a = \cos ka \quad (3.4.5)$$

The cosine function is periodic in 2π , so

$$\cos \alpha a = \cos ka \equiv \cos(ka + n2\pi) \quad (3.4.6)$$

where $n = 0, \pm 1, \pm 2, \pm 3, \dots$ and

$$\alpha a = ka + n2\pi \quad (3.4.7)$$

But $\alpha = (2m_e E / \hbar^2)^{1/2}$, and thus

$$(2m_e / \hbar^2)^{1/2} E^{1/2} = k + n2\pi/a \quad (3.4.8)$$

This indicates that the parabola is repeated periodically with $n2\pi/a$. In other words, the energy is a periodic function of k with the periodicity $2\pi/a$. As mentioned earlier (see Fig. 3.8), discontinuities in energy occur at the boundary of an allowed band when $\cos ka = \pm 1$, i.e., for $k = n\pi/a$ ($n = \pm 1, \pm 2, \pm 3, \dots$).

At these values of k , deviation from the parabolic $E(k)$ curve occurs, indicating that, in a periodic lattice, the electrons behave similar to free particles, except for $k = n\pi/a$. This is shown in Fig. 3.9, which shows the *extended zone representation* of $E(k)$ dependence. This figure also shows the *Brillouin zones*, i.e. the values of k associated with a specific energy band (in the figure, these are identified as 1st zone, 2nd zone, 3rd zone, etc.). Another, more convenient representation is the *reduced-zone representation* (see Fig. 3.10), which can be derived by folding the bands back into the 1st (Brillouin) zone. These $E(k)$ diagrams and other electronic-band-theory terminology, introduced later, are of great importance for describing various properties of semiconductors. As in the description of crystalline structures (see Chapter 2), which distinguishes between the reciprocal (i.e., \mathbf{k} -space) and real lattices, the electronic energy bands can be also represented in (i) \mathbf{k} -space (the reduced-zone representation) and (ii) in real space as a function of space coordinate \mathbf{r} (see Fig. 3.11).

Thus, to summarize briefly, the $E(k)$ relationship according to the periodic potential model is no longer described by a parabolic function (corresponding to free electron model), but it is a function that reveals the presence of various allowed energy bands separated with energy bands that are forbidden, i.e., bands where no permitted k -states can exist. This model is often referred to as *nearly-free electron model*.

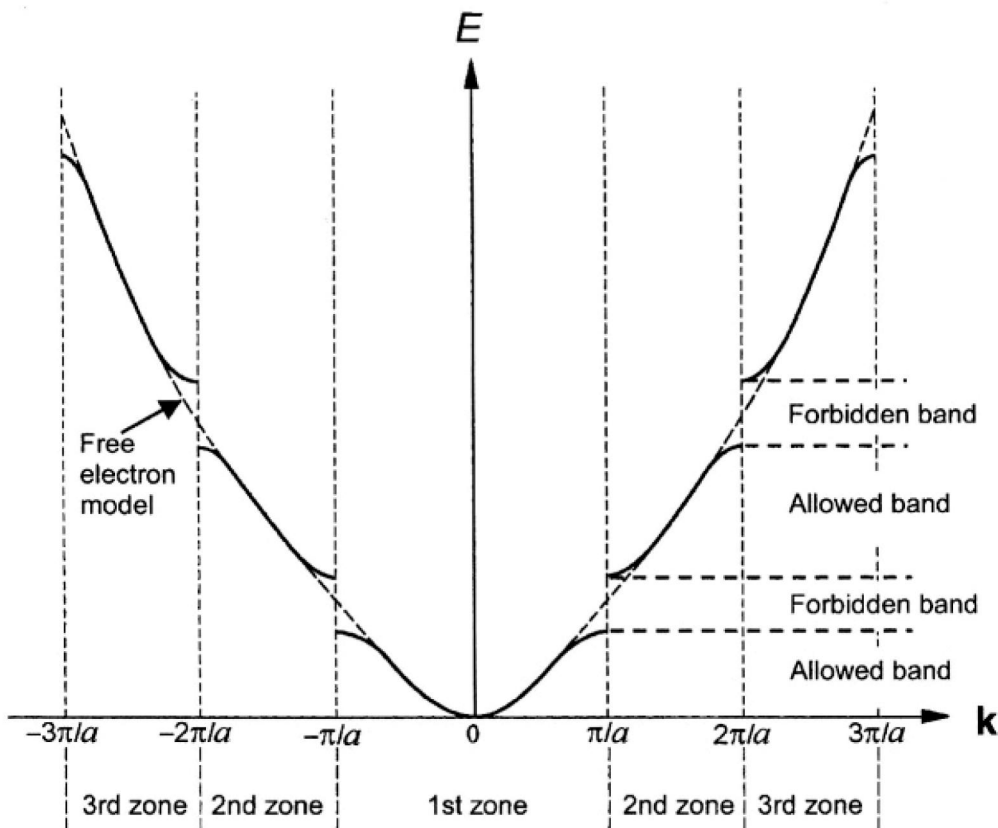


FIGURE 3.9. The extended-zone representation of the nearly-free electron model, showing the modification of the parabolic $E(k)$ dependence for free electrons at the band edges corresponding to $k = n\pi/a$. First three Brillouin zones are also indicated.

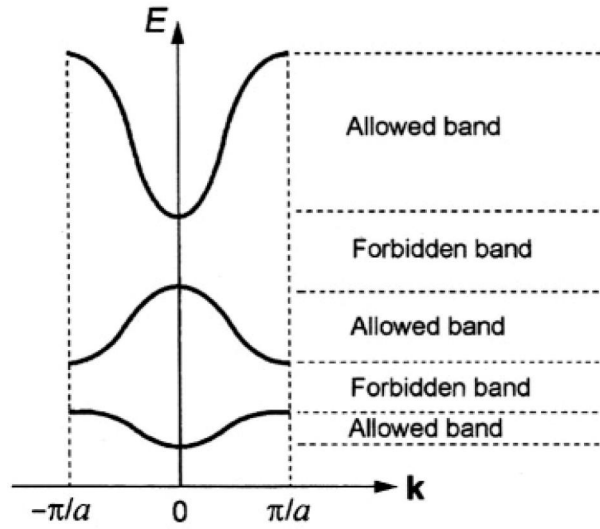
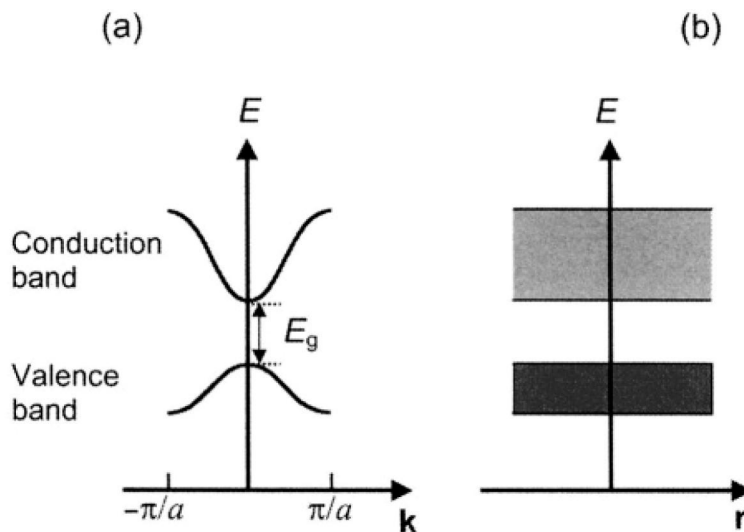


FIGURE 3.10. The reduced-zone representation.

As mentioned earlier, there is also a method that does not require the need of invoking the periodic potential, but instead, the description is based on the analysis of atomic orbitals. In this case, the electrons are considered to be tightly bound to the nuclei, and with the formation of a solid, the electronic wave functions of constituent atoms overlap. These electronic wave functions can be approximated by a linear combination of the atomic wave functions. Such a method is referred to as the *tight-binding approximation* or *linear combination of atomic orbitals* (LCAO). In this method, electron positions in molecular orbitals are approximated by a linear combination of atomic orbitals of the form $\psi = c_1\phi_1 + c_2\phi_2 + c_3\phi_3 + \dots$, where ψ and ϕ correspond to molecular orbital

FIGURE 3.11. Schematic diagram of the electronic energy bands in (a) k -space (the reduced-zone representation) and (b) in real space as a function of space coordinate r .

wave function and atomic orbital wave function, respectively. Thus, the problem of determining the appropriate function for the molecular orbitals is basically reduced to optimizing coefficients in this linear equation. Thus, to summarize briefly, LCAO is essentially an approximation for determining a molecular orbital by using a superposition of atomic orbitals. In employing such an approximation, although the specific atomic orbitals and their number are not initially identified, inclusion of additional atomic orbitals in the linear combination results in the refinement of the approximation. (Note that the main motivation for employing approximate solutions is related to the difficulties in finding the exact solutions to the Schrödinger equation for increasingly complex systems.)

Based on this brief outline of these methods, it follows that the nearly-free electron model is more suited for the calculation of the conduction band states (since electrons in this case are nearly free), whereas the LCAO method is a suitable approximation for the calculation of the valence band states (since the wave functions of valence electrons are comparable with bonding orbitals).

3.5. BRILLOUIN ZONES AND EXAMPLES OF THE ENERGY BAND STRUCTURE FOR SEMICONDUCTORS

The information on the electronic band structure can be illustrated by employing the first Brillouin zone, or as often referred to as Brillouin zone, which is expressed for one period of the reciprocal lattice centered about the origin of the \mathbf{k} -space (see Fig. 3.12). In this case, the edges of the bands are described using the boundaries of a three-dimensional figure in \mathbf{k} -space. In other words, this representation provides a three-dimensional description of the band theory by

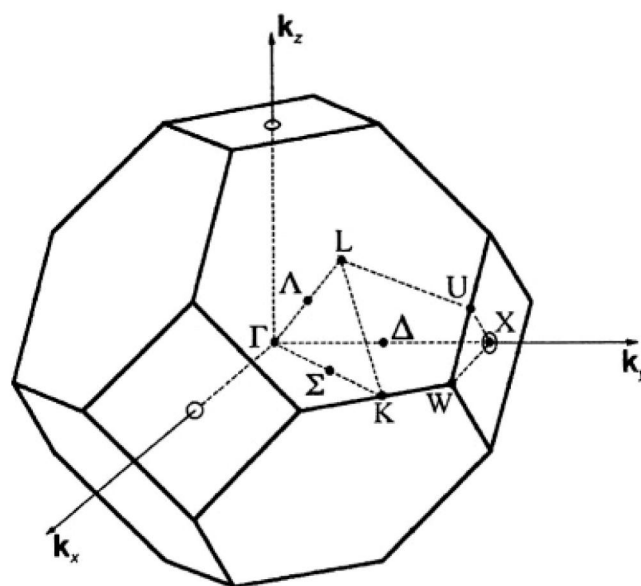


FIGURE 3.12. The first Brillouin zone of the diamond and zincblende-type structures. The important symmetry points and lines are indicated.

using Brillouin zones (i.e., bands of permitted energy). As noted previously, at $k = n\pi/a$, the discontinuities in energy occur at these values of k , and these conditions also correspond to the Bragg reflection rule, suggesting that the electron states with $k = n\pi/a$ can be described as standing waves, i.e., these electrons cannot propagate through the lattice, indicating the presence of the energy gap for that given k . Thus, this essentially also implies that at the boundaries of the Brillouin zones (i.e., $k = n\pi/a$) the electron waves experience reflection from the crystal planes according to the Bragg law, and they cannot propagate through the lattice.

The shape of the Brillouin zone is determined by the crystal lattice geometry and its size depends on the lattice constant. The surface that encloses the occupied states is referred to as the *Fermi surface*. In the description of the Brillouin zone, the standard notation of the points of symmetry inside the zone is in terms of Greek letters (see Fig. 3.12), and it is in terms of Roman letters for the surface. In such a representation, the energy band structure is described along selected crystallographic orientations. Thus, e.g., Γ represents the origin of the reciprocal space (i.e., $k = 0$); Λ represents a direction such as $[111]$ and L denotes the zone end along that direction; Δ represents a direction such as $[100]$ and X denotes the zone end along that direction; and Σ represents a direction such as $[110]$ and K denotes the zone end along that direction. A convenient way of describing the band structure is in terms of $E(\mathbf{k})$ relationship plotted along selected directions in the reduced-zone representation. Such a description of the detailed band structure, derived from theoretical calculations and experimental observations, is of great importance in elucidating the optical and electrical properties of semiconductor

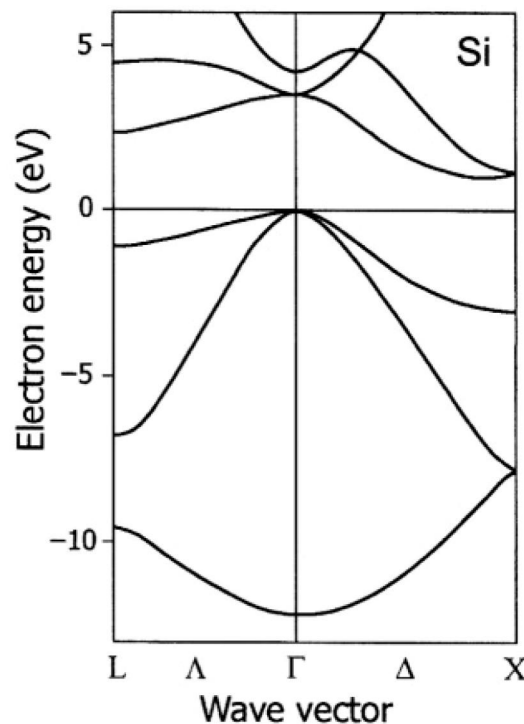


FIGURE 3.13. Electronic energy band structure of Si.

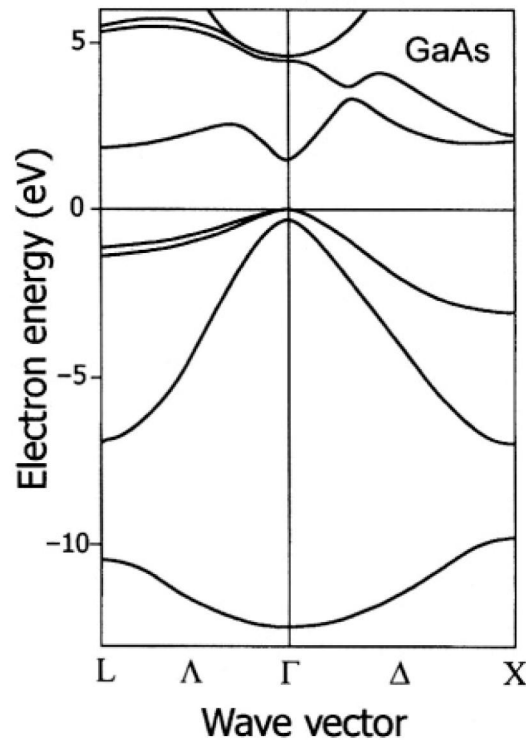


FIGURE 3.14. Electronic energy band structure of GaAs.

and for designing semiconductor devices. The band structures of two most important semiconductors (i.e., Si and GaAs) are shown in Figs. 3.13 and 3.14. Close examination of these diagrams reveals some important facts. As expected from the previous sections, there is a forbidden energy region, which is referred to as the *energy gap*, which separates the *conduction band* from the *valence band*. More specifically, the energy gap (E_g) is defined as the energy separation between the highest valence band maximum and the lowest conduction band minimum. What distinguishes these two important semiconductors (i.e., Si and GaAs) is the location of the lowest conduction band minima in relation to the highest valence band maxima. In the case of GaAs (see Fig. 3.14), both the highest valence band maximum and the lowest conduction band minimum are at the same Γ point; such a material is referred to as *direct energy-gap semiconductor* (this fact is of great importance, e.g., in optical processes such as electronic transitions between the valence and conduction bands, see Chapter 4). In the case of Si (see Fig. 3.13), the highest valence band maximum is at Γ point, but the lowest conduction band minimum is in the Δ or [100] direction near the first Brillouin zone boundary (i.e., X point). In such a case, when the highest valence band maximum and the lowest conduction band minimum are not at the same point in \mathbf{k} -space, a semiconductor is referred to as *indirect energy-gap material* (again, this fact is of great importance in optical processes, see Chapter 4). An additional important feature observed in the band structure of GaAs is the presence of the conduction band minima at L and X points, which are referred to as L-valley and X-valley and which have a major effect on high-field transport properties.

3.6. THE EFFECTIVE MASS

The concept of effective mass is an important consideration for the description of the dynamics of the movement of electrons in a crystal. In semiconductors, the presence of the periodic crystal potential modifies the electron properties, resulting in the electron mass that is different from the free electron mass. The *effective mass* (denoted as m^*) is typically referred to the experimentally determined value of the electron mass, and it is usually given in terms of the free electron mass, i.e., m^*/m_0 . For different semiconductors, this ratio can be slightly greater or less than unity. In order to derive an expression for the effective mass in terms of the electronic band structure parameters, we recall that, for a wave packet, the group velocity ($v_g = d\omega/dk$) can be expressed as (note that $\omega = E/\hbar$)

$$v_g = \frac{1}{\hbar} \frac{dE}{dk} \quad (3.6.1)$$

and the acceleration a can be expressed as

$$a = \frac{dv_g}{dt} = \frac{1}{\hbar} \frac{d^2E}{dk^2} \frac{dk}{dt} \quad (3.6.2)$$

dE/dk is known, and dk/dt can be evaluated from the expression $p = \hbar k$

$$\frac{dp}{dt} = \hbar \frac{dk}{dt} \quad (3.6.3)$$

Thus,

$$a = \frac{1}{\hbar^2} \frac{d^2E}{dk^2} \frac{dp}{dt} = \frac{1}{\hbar^2} \frac{d^2E}{dk^2} F \quad (3.6.4)$$

where F is the force acting on an electron. By comparing Eq. (3.6.4) with an equation for acceleration, i.e., $a = F/m$, we can write

$$\frac{1}{m^*} = \frac{1}{\hbar^2} \frac{d^2E}{dk^2} \quad (3.6.5)$$

or

$$m^* = \hbar^2 \left(\frac{d^2E}{dk^2} \right)^{-1} \quad (3.6.6)$$

In other words, the electron effective mass is inversely proportional to the curvature of an electron band $E(k)$. This implies that the carrier effective mass can

be determined from the electronic band structure, or $E(k)$ relationship. For example, the examination of the extended zone representation of $E(k)$ diagram (see Fig. 3.9) indicates that (i) the effective mass is positive near the bottom of the bands, and it is negative near the top of the bands and (ii) the effective mass near both the top and bottom of an energy band is energy-independent, i.e., constant. Electrons having the negative mass near the top of the band are referred to as an “electron hole”. However, in the presence of an electric field, a negative mass with a negative charge is equivalent to a positive mass with a positive charge (i.e., $-ma = -e\mathcal{E}$ is equivalent to $ma = e\mathcal{E}$), and thus, the *holes* are described as carriers having a positive mass and a positive charge. This implies that in the presence of an electric field, electrons at the bottom of the conduction band and holes at the top of the valence band travel in the opposite directions in real space. As it is discussed in the subsequent chapters, the electrical transport in a nearly filled band (i.e., valence band) in terms of the motion of holes is of great importance in describing semiconductor properties.

The inverse dependence of the effective mass on the $E(k)$ curvature indicates that the greater curvature implies smaller effective mass, and vice versa. In addition, since properties of the material depend on crystallographic directions (i.e., anisotropy), the effective mass may differ in each direction. Considering the case of GaAs (see Fig. 3.14), effective masses for electrons in the conduction band differ, e.g., between the central Γ -valley (with greater curvature) and L-valley (with smaller curvature). These masses are $0.067m_0$ and $0.35m_0$ for Γ -valley and L-valley, respectively. The energy at the lowest minimum at L is only about 0.3 eV higher than that at Γ . This difference in the effective mass has an important influence on transport properties of the semiconductors at sufficiently high fields

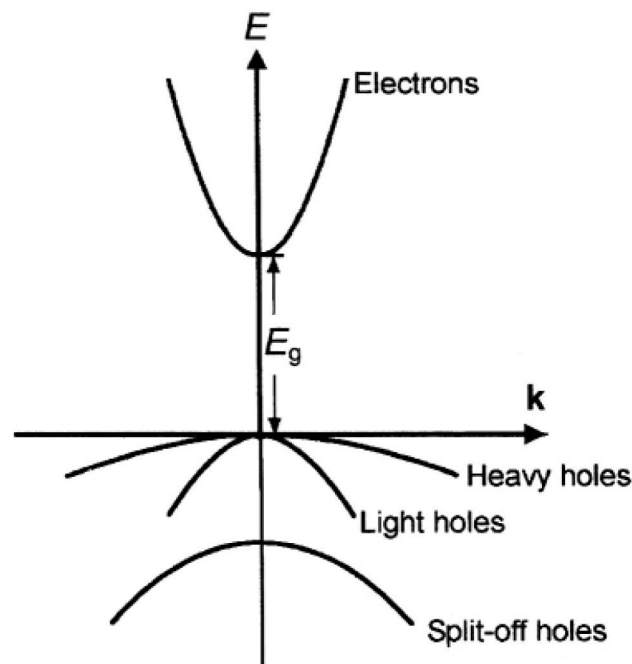


FIGURE 3.15. $E(k)$ diagram illustrating various hole masses, i.e., light holes, heavy holes, and split-off holes.

TABLE 3.1. Effective masses of electrons, m_e^* , holes (light, m_{lh}^* , and heavy, m_{hh}^*) and split-off holes m_{soh}^* in units of m_0 (i.e., the free electron rest mass) for selected direct energy-gap semiconductors

	m_e^*/m_0	m_{lh}^*/m_0	m_{hh}^*/m_0	m_{soh}^*/m_0
InAs	0.026	0.025	0.41	0.08
InP	0.073	0.078	0.40	0.15
GaAs	0.067	0.082	0.50	0.15

(greater than about 3 kV cm^{-1}), since some of the conduction electrons can be transferred into the L-valley with its greater effective mass and can subsequently affect the carrier mobility and result in reduced drift velocity. In the valence band (see Fig. 3.15), similar considerations of the inverse dependence between the effective mass and energy band curvature indicate that there are the *light-hole* band (larger curvature) and *heavy-hole* band (smaller curvature). In addition, in the valence band, the *split-off* band at a slightly lower energy (due to spin–orbit interactions) is also present (Fig. 3.15). Tables 3.1 and 3.2 list the effective masses of both the electrons and holes in selected direct and indirect energy-gap semiconductors. In Chapter 4, the concept of the density of states in the conduction and valence bands will be introduced. The calculations of these quantities involve the appropriate choice of the effective mass, since the band-structure representation of a given band typically invokes different effective masses; thus, this requires employing a combination of the effective masses. Since holes typically occupy the light-hole and heavy-hole bands only, for calculations of the density of states in the valence band, the effective mass of the valence band is expressed as

$$m_{hd}^* = \left[(m_{lh}^*)^{3/2} + (m_{hh}^*)^{3/2} \right]^{2/3} \quad (3.6.7)$$

For the conduction band, the effective mass for density of states calculations is

$$m_{ed}^* = N_{bm}^{2/3} (m_x^* m_y^* m_z^*)^{1/3} \quad (3.6.8)$$

where N_{bm} is the number of equivalent band minima in the conduction band, and m_x^* , m_y^* , and m_z^* are the effective masses along the principal axes of the ellipsoidal energy surface. (Note that in semiconductors, such as Si, constant-energy surfaces

TABLE 3.2. Effective masses of electrons (longitudinal, m_l^* , and transverse, m_t^*), holes (light, m_{lh}^* , and heavy, m_{hh}^*) and split-off holes m_{soh}^* in units of m_0 (i.e., the free electron rest mass) for indirect energy-gap semiconductors Ge and Si

	m_l^*/m_0	m_t^*/m_0	m_{lh}^*/m_0	m_{hh}^*/m_0	m_{soh}^*/m_0
Ge	1.64	0.082	0.04	0.35	0.08
Si	0.98	0.19	0.16	0.54	0.23

in the conduction band are ellipsoids, and thus, the effective masses are characterized by the curvature along the long axis as longitudinal m_l^* and transverse m_t^* masses.) Thus, for Si, for calculations of the density of states in the conduction band, the effective mass of the electron is expressed as

$$m_{ed}^* = 6^{2/3} (m_l^* m_t^{*2})^{1/3} \quad (3.6.9)$$

The effective mass for conductivity (or mobility) calculations is derived from their inverse relationship, i.e., conductivity is proportional to the sum of the inverse of the individual masses, and thus the conductivity effective mass is

$$m_{ec}^* = 3 \left(\frac{1}{m_l} + \frac{2}{m_t} \right)^{-1} \quad (3.6.10)$$

One of the methods that is often used for measuring effective mass is the *cyclotron resonance* (i.e., resonant absorption of electromagnetic waves) experiment. In this case, an interaction of electromagnetic waves with charge carriers results in the resonant absorption of the electromagnetic wave in the presence of an applied magnetic field that causes the carrier to vibrate at the same frequency as the frequency of the applied electric field. The effective mass can be derived using equation $m^* = eB/\omega_c$, where B is the magnetic field corresponding to the maximum absorption of the electromagnetic wave having frequency equal to cyclotron frequency ω_c .

3.7. CLASSIFICATION OF SOLIDS ACCORDING TO THE BAND THEORY

The electronic band structure of solids can explain the distinction between metals, semiconductors, and insulators (see Fig. 3.16). First, it should be emphasized that empty energy bands, which do not contain electrons, do not contribute to the electrical conductivity of a material; whereas, although completely filled bands contain electrons, they do not contribute to the electrical conductivity, since the carriers are unable to gain energy (when an electric field is applied) due to the fact that all the energy levels are occupied. In comparison, the partially filled bands contain both the electrons and unoccupied energy levels at higher energies. The latter allow carriers to gain energy in the presence of an applied electric field, and thus, carriers in a partially filled band can contribute to the electrical conductivity of the material. In the case of metals, the electronic band structure results in incomplete filling of the highest occupied energy band. (Note that the metallic behavior may also be a result of an energy overlap between filled and empty bands, not shown in the figure.) The magnitude of the energy gap separating the highest filled (or valence) band and the lowest empty (or conduction) band distinguishes a semiconductor from an insulator. In the case of an intrinsic semiconductor, the highest filled (valence) band is separated

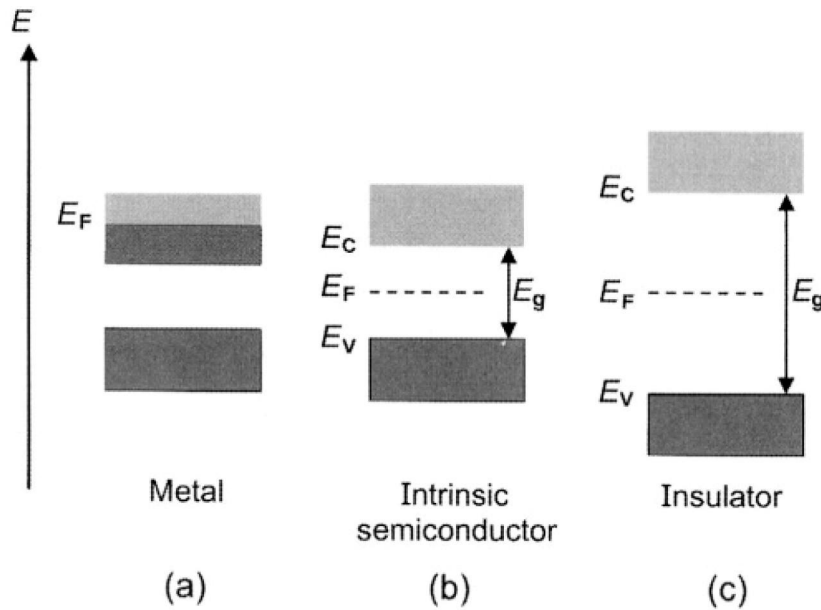


FIGURE 3.16. Schematic illustration of typical band diagrams for (a) a metal, (b) an intrinsic semiconductor ($T = 0$ K), and (c) an insulator. Dashed lines represent the Fermi level (for details, see Chapter 4). Darker regions represent filled bands, whereas lighter regions correspond to the empty bands.

from the lowest empty (conduction) band by a relatively narrow forbidden energy gap, and at $T = 0$ K there are no electrons in the conduction band. However, in semiconductors, the energy gap (E_g) is sufficiently small, so that at room temperature the electrons from the top of the valence band are thermally excited to the conduction band, where they can contribute to the carrier transport in a material. In insulators, the energy gap is so much greater (as compared to semiconductors) that at room temperature the probability of thermal excitation of an electron from the valence band to the conduction band is very low. In Fig. 3.16, the *Fermi energy*, or *Fermi level* (E_F) defines the reference energy for the probability of occupation of electron states (for more details, see Chapter 4). Thus, in metals E_F is located within a partially filled allowed band, as shown in Fig. 3.16, whereas in semiconductors and insulators E_F is positioned within the forbidden band.

At this juncture, it is also important to define an additional parameter that is useful in the description of semiconductors (see Fig. 3.17). This is the *electron affinity* ($e\chi$), defined as the energy difference between the vacuum level (i.e., the energy of a free electron) and the bottom of the conduction band. Another important parameter is the *work function* (often denoted as $e\Phi$), which is the energy difference between the vacuum level and the Fermi level. It should be noted that the electron affinity ($e\chi$) is a constant for a given semiconductor, whereas the work function $e\Phi$ depends on the *doping*, which affects the Fermi level position. (Doping is the process of putting impurities into the lattice to produce materials and devices with desired properties, see Chapter 4.)

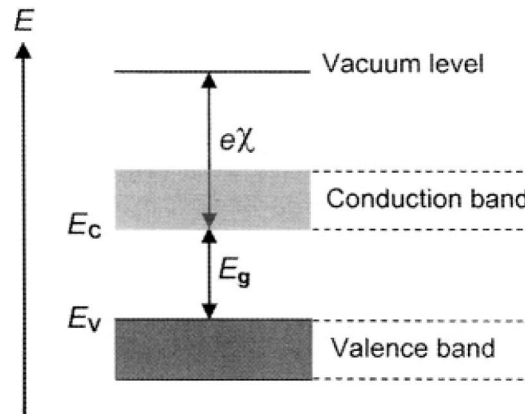


FIGURE 3.17. Schematic illustration of a typical band diagram indicating some important definitions for the description of a semiconductor; this includes the electron affinity ($e\chi$), i.e., the energy difference between the vacuum level and the bottom of the conduction band.

3.8. SUMMARY

The electronic properties in semiconductors can be described by the quantum theory of the electronic energy band structure of solids. In one of the approaches for the description of electronic band structure, with the formation of a solid, the electronic wave functions of constituent atoms overlap, and the application of the Pauli exclusion principle leads to the splitting of the discrete energy levels of the isolated atoms into bands of allowed electron levels separated by forbidden energy gaps.

According to the *Kronig–Penney model* (i.e., the electron in a periodic crystal potential), the electrons, moving in a periodically varying potential field, may possess energies within certain energy bands only, i.e., resulting in the presence of allowed and forbidden bands of energy for electrons moving through a periodic potential. The presence of the periodic crystal potential also results in the electron mass (i.e., the *effective mass*) that is different from the free electron mass. The electron effective mass is inversely proportional to the curvature of an electron band $E(k)$. Electrons having the negative mass near the top of the band are referred to as an “electron hole,” or equivalently as the *holes*, i.e., carriers having a positive mass and a positive charge. The inverse dependence of the effective mass on the energy band curvature indicates that there are light-hole band (larger curvature) and heavy-hole band (smaller curvature) in the valence band. Also, since properties of the material depend on crystallographic directions (i.e., anisotropy), the effective mass may differ in each direction.

The electronic properties of a solid are determined by the electron occupation of the highest energy bands, i.e., the valence and conduction bands, which are separated by a *fundamental energy gap*, E_g . Thus, the quantum-mechanical description of the electronic band structure provides a clear distinction between conductors, semiconductors and insulators. In conductors, the electronic band structure results in incomplete filling of the highest occupied energy band,