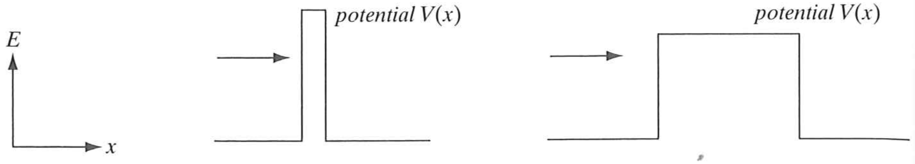
Question 4

Consider quantum mechanical particles incident from the left having well-defined energy as indicated by the vertical positions of the arrows, in the two systems shown below. Will the probability of being reflected be greater for the incident particle in System 1 than for the incident particle in System 2? Circle one choice below.

yes / no / not enough information provided



System 1
(narrow high potential barrier)

System 2
(wide low potential barrier)

Question 5

Suppose five precise measurements were made on a particle in rapid succession, such that the time evolution of the particle wave function *between* measurements could be neglected, in the following order: (1) position, (2) momentum, (3) momentum, (4) position, (5) momentum. If the results of the first two measurements were x_o and p_o, respectively, what would be the results of the next three measurements (circle one each)?

measurement (3): momentum **p_o / unknown**

measurement (4): position **x_o / unknown**

measurement (5): momentum **p_o / unknown**

Question 6

If the photoelectric effect were governed by classical physics rather than quantum mechanics, what would be result of the following experiments:

(a) By changing the intensity of the incident radiation, what would happen to the energy and number of ejected electrons?

(b) How about changing the frequency of the light?

Chapter 3

Energy Bands and Charge Carriers in Semiconductors

OBJECTIVES

1. Understand conduction and valence energy bands, and how bandgaps are formed
2. Appreciate the idea of doping in semiconductors
3. Use the density of states and Fermi Dirac statistics to calculate carrier concentrations
4. Calculate drift currents in an electric field in terms of carrier mobility, and how mobility is affected by scattering
5. Discuss the idea of "effective" masses

In this chapter we begin to discuss the specific mechanisms by which current flows in a solid. In examining these mechanisms we shall learn why some materials are good conductors of electric current, whereas others are poor conductors. We shall see how the conductivity of a semiconductor can be varied by changing the temperature or the number of impurities. These fundamental concepts of charge transport form the basis for later discussions of solid state device behavior.

3.1
BONDING FORCES AND ENERGY BANDS IN SOLIDS

In Chapter 2 we found that electrons are restricted to sets of discrete energy levels within atoms. Large gaps exist in the energy scale in which no energy states are available. In a similar fashion, electrons in solids are restricted to certain energies and are not allowed at other energies. The basic difference between the case of an electron in a solid and that of an electron in an isolated atom is that in the solid the electron has a *range,* or *band,* of available energies. The discrete energy levels of the isolated atom spread into bands of energies in the solid because in the solid the wave functions of electrons in neighboring atoms overlap, and an electron is not necessarily localized at a particular atom. Thus, for example, an electron in the outer orbit of one atom feels the influence of neighboring atoms, and its overall wave function is altered. Naturally, this influence affects the potential energy term and the boundary conditions in the Schrödinger equation, and we would expect to obtain different energies in the solution. Usually, the influence of neighboring atoms on the energy levels of a particular atom can be treated as a small perturbation, giving rise to shifting and splitting of energy states into energy bands.

3.1.1 Bonding Forces in Solids

The interaction of electrons in neighboring atoms of a solid serves the very important function of holding the crystal together. For example, alkali halides such as NaCl are typified by *ionic bonding*. In the NaCl lattice, each Na atom is surrounded by six nearest neighbor Cl atoms, and vice versa. Four of the nearest neighbors are evident in the two-dimensional representation shown in Fig. 3–1a. The electronic structure of Na ($Z = 11$) is [Ne] $3s^1$, and Cl ($Z = 17$) has the structure [Ne]$3s^2 3p^5$. In the lattice each Na atom gives up its outer $3s$ electron to a Cl atom, so that the crystal is made up of ions with the electronic structures of the inert atoms Ne and Ar (Ar has the electronic structure [Ne]$3s^2 3p^6$). However, the ions have net electric charges after the electron exchange. The Na^+ ion has a net positive charge, having lost an electron, and the Cl^- ion has a net negative charge, having gained an electron.

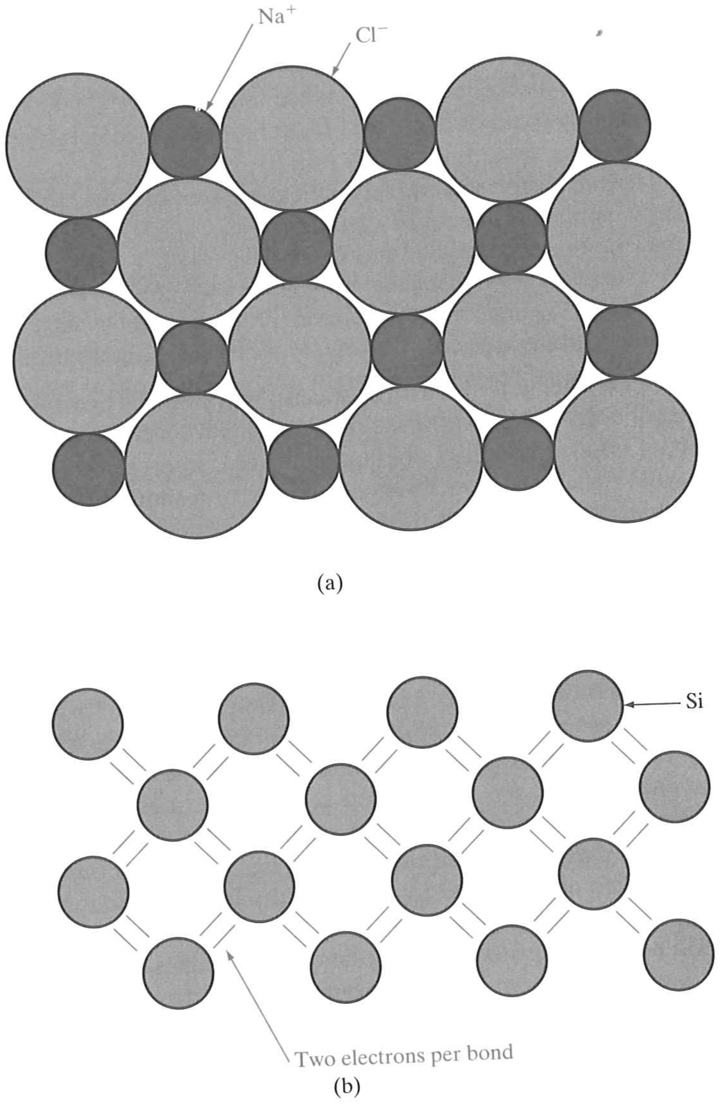


Figure 3–1
Different types of chemical bonding in solids: (a) an example of ionic bonding in NaCl; (b) covalent bonding in the Si crystal, viewed along a <100> direction (see also Figs. 1–8 and 1–9).

Each Na^+ ion exerts an electrostatic attractive force upon its six Cl^- neighbors, and vice versa. These coulombic forces pull the lattice together until a balance is reached with repulsive forces. A reasonably accurate calculation of the atomic spacing can be made by considering the ions as hard spheres being attracted together (Example 1-1).

An important observation in the NaCl structure is that all electrons are tightly bound to atoms. Once the electron exchanges have been made between the Na and Cl atoms to form the Na^+ and Cl^- ions, the outer orbits of all atoms are completely filled. Since the ions have the closed-shell configurations of the inert atoms Ne and Ar, there are no loosely bound electrons to participate in current flow; as a result, NaCl is a good insulator.

In a metal atom the outer electronic shell is only partially filled, usually by no more than three electrons. We have already noted that the alkali metals (e.g., Na) have only one electron in the outer orbit. This electron is loosely bound and is given up easily in ion formation. This accounts for the great chemical activity in the alkali metals, as well as for their high electrical conductivity. In the metal the outer electron of each alkali atom is contributed to the crystal as a whole, so that the solid is made up of ions with closed shells immersed in a sea of free electrons. The forces holding the lattice together arise from an interaction between the positive ion cores and the surrounding free electrons. This is one type of *metallic bonding*. Obviously, there are complicated differences in the bonding forces for various metals, as evidenced by the wide range of melting temperatures (234 K for Hg, 3643 K for W). However, the metals have the sea of electrons in common, and these electrons are free to move about the crystal under the influence of an electric field.

A third type of bonding is exhibited by the diamond lattice semiconductors. We recall that each atom in the Ge, Si, or C diamond lattice is surrounded by four nearest neighbors, each with four electrons in the outer orbit. In these crystals each atom shares its valence electrons with its four neighbors (Fig. 3-1b). Bonding between nearest neighbor atoms is illustrated in the diamond lattice diagram of Fig. 1-9. The bonding forces arise from a quantum mechanical interaction between the shared electrons. This is known as *covalent bonding;* each electron pair constitutes a covalent bond. In the sharing process it is no longer relevant to ask which electron belongs to a particular atom—both belong to the bond. The two electrons are indistinguishable, except that they must have opposite spin to satisfy the Pauli exclusion principle. Covalent bonding is also found in certain molecules, such as H_2.

As in the case of the ionic crystals, no free electrons are available to the lattice in the covalent diamond structure of Fig. 3–1b. By this reasoning Ge and Si should also be insulators. However, we have pictured an idealized lattice at 0 K in this figure. As we shall see in subsequent sections, an electron can be thermally or optically excited out of a covalent bond and thereby become free to participate in conduction. This is an important feature of semiconductors.

Compound semiconductors such as GaAs have mixed bonding, in which both ionic and covalent bonding forces participate. Some ionic bonding is to be expected in a crystal such as GaAs because of the difference in placement of the Ga and As atoms in the periodic table. The ionic character of the bonding becomes more important as the atoms of the compound become further separated in the periodic table, as in the II–VI compounds. Such electronic structure, and specifically the idea that the outermost valence shell is *complete* if it has a stable set of eight electrons (Ne, Ar, Kr), is the basis of most of chemistry and many of the semiconducting properties.

3.1.2 Energy Bands

As isolated atoms are brought together to form a solid, various interactions occur between neighboring atoms, including those described in the preceding section. The forces of attraction and repulsion between atoms will find a balance at the proper interatomic spacing for the crystal. In the process, important changes occur in the electron energy level configurations, and these changes result in the varied electrical properties of solids.

In Fig. 2–8, we showed the orbital model of a Si atom, along with the energy levels of the various electrons in the coulombic potential well of the nucleus. Let us focus on the outermost shell or valence shell, $\mathbf{n} = 3$, where two $3s$ and two $3p$ electrons interact to form the four "hybridized" sp^3 electrons when the atoms are brought close together. In Fig. 3–2, we schematically show the coulombic potential wells of two atoms close to each other, along with the wave functions of two electrons centered on the two nuclei. By solving the Schrödinger equation for such an interacting system, we find that the composite two-electron wave functions are *linear combinations* of the individual *atomic orbitals* (LCAO). The odd or antisymmetric combination is called the antibonding orbital, while the even or symmetric combination is the bonding orbital. It can be seen that the bonding orbital has a higher value of the wave function (and therefore the electron probability density) than the antibonding state in the region between the two nuclei. This corresponds to the covalent bond between the atoms.

To determine the energy levels of the bonding and the antibonding states, it is important to recognize that in the region between the two nuclei the coulombic potential energy V(r) is lowered (solid line in Fig. 3–2) compared to isolated atoms (dashed lines). It is easy to see why the potential energy would be lowered in this region, because an electron here would be attracted by two nuclei, rather than just one. For the bonding state the electron probability density is higher in this region of lowered potential energy than for the antibonding state. As a result, the original isolated atomic energy level would be split into two, a lower bonding energy level and a higher antibonding level. It is the lowering of the energy of the bonding state that gives rise to cohesion of the crystal. For even smaller interatomic spacings, the energy of the crystal goes up because of repulsion between the nuclei, and other electronic interactions. Since the probability density is given by the square of the wave function, if the entire wave function is multiplied by −1, it does not lead to a different LCAO. The important point to note in this discussion is that the number of distinct LCAO, and the number of distinct energy levels, depends on the number of atoms that are brought together. The lowest energy level corresponds to the totally symmetric LCAO, the highest corresponds to the totally antisymmetric case, and the other combinations lead to energy levels in between.

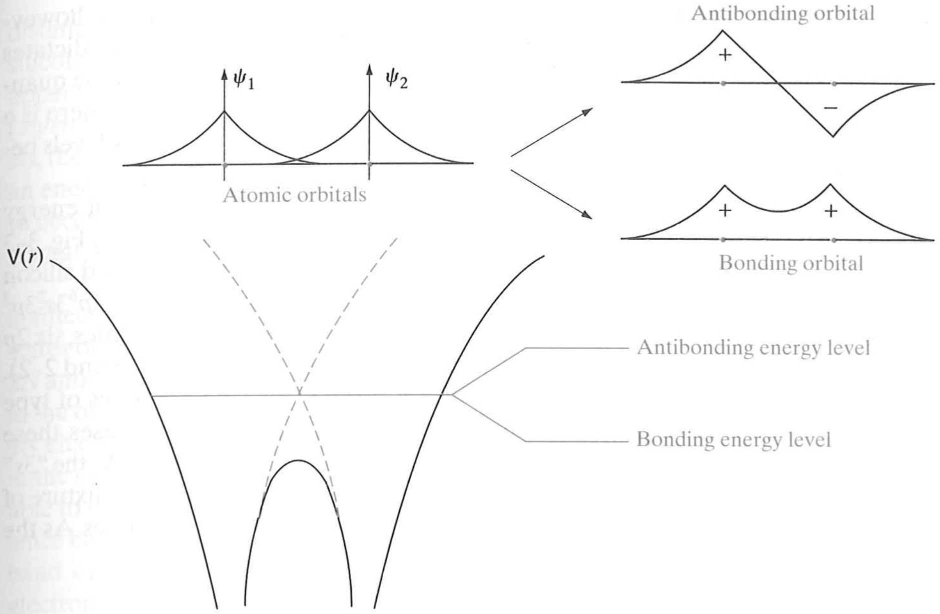


Figure 3–2
Linear combinations of atomic orbitals (LCAO): The LCAO when 2 atoms are brought together leads to 2 distinct "normal" modes—a higher energy antibonding orbital, and a lower energy bonding orbital. Note that the electron probability density is high in the region between the ion cores (covalent "bond"), leading to lowering of the bonding energy level and the cohesion of the crystal. If instead of 2 atoms, one brings together N atoms, there will be N distinct LCAO, and N closely spaced energy levels in a band.

Qualitatively, we can see that as atoms are brought together, the application of the Pauli exclusion principle becomes important. When two atoms are completely isolated from each other so that there is no interaction of electron wave functions between them, they can have identical electronic structures. As the spacing between the two atoms becomes smaller, however, electron wave functions begin to overlap. The exclusion principle dictates that no two electrons in a given interacting system may have the same quantum state; thus there must be at most one electron per level after there is a splitting of the discrete energy levels of the isolated atoms into new levels belonging to the pair rather than to individual atoms.

In a solid, many atoms are brought together, so that the split energy levels form essentially continuous *bands* of energies. As an example, Fig. 3-3 illustrates the imaginary formation of a silicon crystal from isolated silicon atoms. Each isolated silicon atom has an electronic structure $1s^22s^22p^63s^23p^2$ in the ground state. Each atom has available two $1s$ states, two $2s$ states, six $2p$ states, two $3s$ states, six $3p$ states, and higher states (see Tables 2-1 and 2-2). If we consider N atoms, there will be $2N$, $2N$, $6N$, $2N$, and $6N$ states of type $1s$, $2s$, $2p$, $3s$, and $3p$, respectively. As the interatomic spacing decreases, these energy levels split into bands, beginning with the outer ($\mathbf{n} = 3$) shell. As the "$3s$" and "$3p$" bands grow, they merge into a single band composed of a mixture of energy levels. This band of "$3s$-$3p$" levels contains $8N$ available states. As the

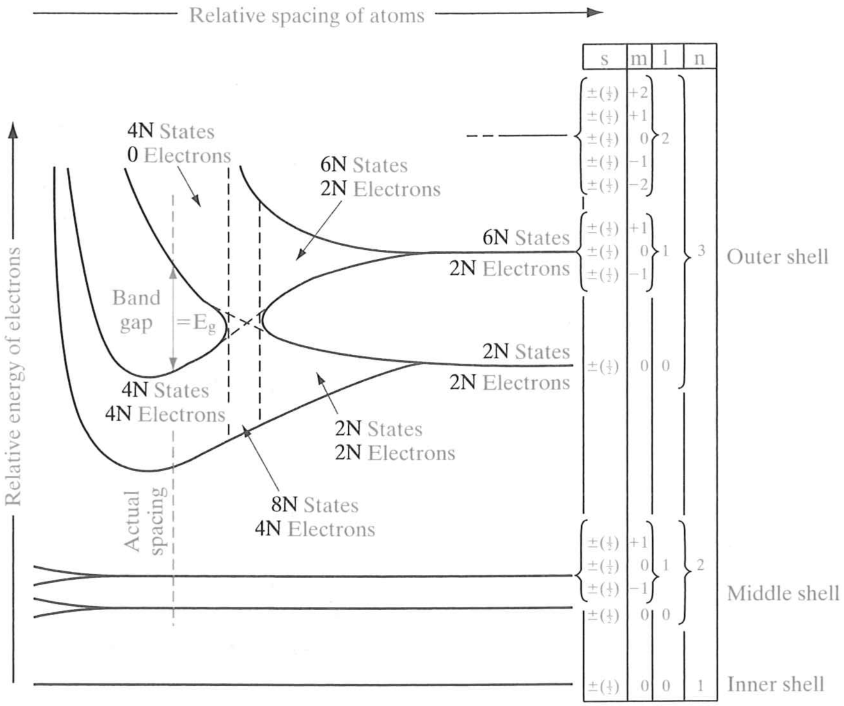


Figure 3–3

Energy levels in Si as a function of interatomic spacing. The core levels ($\mathbf{n} = 1,2$) in Si are completely filled with electrons. At the actual atomic spacing of the crystal, the $2N$ electrons in the $3s$ subshell and the $2N$ electrons in the $3p$ subshell undergo sp^3 hybridization, and all end up in the lower $4N$ states (valence band), while the higher-lying $4N$ states (conduction band) are empty, separated by a band gap.

distance between atoms approaches the equilibrium interatomic spacing of silicon, this band splits into two bands separated by an *energy gap* E_g. The upper band (called the *conduction band*) contains $4N$ states, as does the lower (*valence*) band. Thus, apart from the low-lying and tightly bound "core" levels, the silicon crystal has two bands of available energy levels separated by an energy gap E_g wide, which contains no allowed energy levels for electrons to occupy. This gap is sometimes called a "forbidden band," since in a perfect crystal it contains no electron energy states.

We should pause at this point and count electrons. The lower "$1s$" band is filled with the $2N$ electrons which originally resided in the collective $1s$ states of the isolated atoms. Similarly, the $2s$ band and the $2p$ bands will have $2N$ and $6N$ electrons in them, respectively. However, there were $4N$ electrons in the original isolated $\mathbf{n} = 3$ shells ($2N$ in $3s$ states and $2N$ in $3p$ states). These $4N$ electrons must occupy states in the valence band or the conduction band in the crystal. At 0 K the electrons will occupy the lowest energy states available to them. In the case of the Si crystal, there are exactly $4N$ states in the valence band available to the $4N$ electrons. Thus at 0 K, *every* state in the valence band will be filled, while the conduction band will be completely empty of electrons. As we shall see, this arrangement of completely filled and empty energy bands has an important effect on the electrical conductivity of the solid.

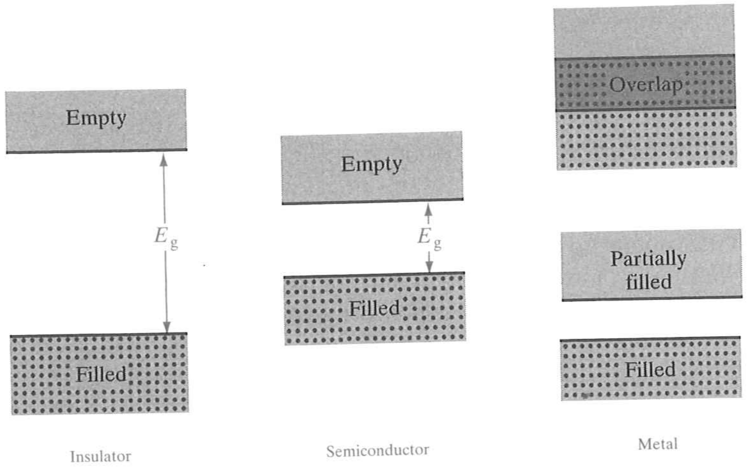
3.1.3 Metals, Semiconductors, and Insulators

Every solid has its own characteristic energy band structure. This variation in band structure is responsible for the wide range of electrical characteristics observed in various materials. The silicon band structure of Fig. 3-3, for example, can give a good picture of why silicon in the diamond lattice is a good insulator. To reach such a conclusion, we must consider the properties of completely filled and completely empty energy bands in the current conduction process.

Before discussing the mechanisms of current flow in solids further, we can observe here that for electrons to experience acceleration in an applied electric field, they must be able to move into new energy states. This implies there must be empty states (allowed energy states which are not already occupied by electrons) available to the electrons. For example, if relatively few electrons reside in an otherwise empty band, ample unoccupied states are available into which the electrons can move. On the other hand, the silicon band structure is such that the valence band is completely filled with electrons at 0 K and the conduction band is empty. There can be no charge transport within the valence band, since no empty states are available into which electrons can move. There are no electrons in the conduction band, so no charge transport can take place there either. Thus silicon has a high resistivity typical of insulators.

Semiconductor materials at 0 K have basically the same structure as insulators—a filled valence band separated from an empty conduction band by a band gap containing no allowed energy states (Fig. 3-4). The difference lies in the size of the band gap E_g, which is much smaller in semiconductors

Figure 3–4
Typical band structures at 0 K.



than in insulators. For example, the semiconductor Si has a band gap of about 1.1 eV compared with 5 eV for diamond. The relatively small band gaps of semiconductors (Appendix III) allow for excitation of electrons from the lower (valence) band to the upper (conduction) band by reasonable amounts of thermal or optical energy. For example, at room temperature a semiconductor with a 1-eV band gap will have a significant number of electrons excited thermally across the energy gap into the conduction band, whereas an insulator with $E_g = 10$ eV will have a negligible number of such excitations. Thus an important difference between semiconductors and insulators is that the number of electrons available for conduction can be increased greatly in semiconductors by thermal or optical energy.

In metals the bands either overlap or are only partially filled. Thus electrons and empty energy states are intermixed within the bands so that electrons can move freely under the influence of an electric field. As expected from the metallic band structures of Fig. 3–4, metals have a high electrical conductivity.

3.1.4 Direct and Indirect Semiconductors

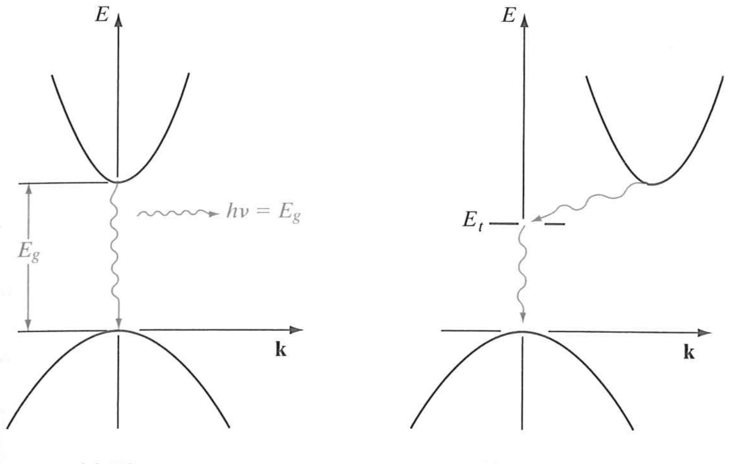
The "thought experiment" of Section 3.1.2, in which isolated atoms were brought together to form a solid, is useful in pointing out the existence of energy bands and some of their properties. Other techniques are generally used, however, when quantitative calculations are made of band structures. In a typical calculation, a single electron is assumed to travel through a perfectly periodic lattice. The wave function of the electron is assumed to be in the form of a plane wave[1] moving, for example, in the x- direction with propagation constant $\mathbf{k}$, also called a *wave vector*. The space-dependent wave function for the electron is

$$\psi_{\mathbf{k}}(x) = U(\mathbf{k}_x, x)e^{j\mathbf{k}_x x} \tag{3-1}$$

where the function $U(\mathbf{k}_x, x)$ modulates the wave function according to the periodicity of the lattice.

---

[1] Discussions of plane waves are available in most sophomore physics texts or in introductory electromagnetics texts.



(a) Direct (b) Indirect

Figure 3–5
Direct and indirect electron transitions in semiconductors: (a) direct transition with accompanying photon emission; (b) indirect transition via a defect level.

In such a calculation, allowed values of energy can be plotted vs. the propagation constant $\mathbf{k}$. Since the periodicity of most lattices is different in various directions, the $(E, \mathbf{k})$ diagram must be plotted for the various crystal directions, and the full relationship between E and $\mathbf{k}$ is a complex surface which should be visualized in three dimensions.

The band structure of GaAs has a minimum in the conduction band and a maximum in the valence band for the same $\mathbf{k}$ value ($\mathbf{k} = 0$). On the other hand, Si has its valence band maximum at a different value of $\mathbf{k}$ than its conduction band minimum. Thus an electron making a smallest-energy transition from the conduction band to the valence band in GaAs can do so without a change in $\mathbf{k}$ value; on the other hand, a transition from the minimum point in the Si conduction band to the maximum point of the valence band requires some change in $\mathbf{k}$. Thus there are two classes of semiconductor energy bands; *direct* and *indirect* (Fig. 3-5). We can show that an indirect transition, involving a change in $\mathbf{k}$, requires a change of momentum for the electron.

EXAMPLE 3–1

Assuming that U is constant in Eq. (3-1) for an essentially free electron, show that the x-component of the electron momentum in the crystal is given by $\langle p_x \rangle = \hbar \mathbf{k}_x$.

SOLUTION From Eq. (3-1)

$$\psi_{\mathbf{k}}(x) = Ue^{j\mathbf{k}_x x}$$

Using Eq. (2-21b) and the momentum operator,

$$\langle \mathrm{p}_x \rangle = \frac{\displaystyle\int_{-\infty}^{\infty} U^2 e^{-j\mathbf{k}_x x} \frac{\hbar}{j} \frac{\partial}{\partial x} (e^{j\mathbf{k}_x x})\, dx}{\displaystyle\int_{-\infty}^{\infty} U^2\, dx}$$

$$= \frac{\hbar \mathbf{k}_x \displaystyle\int_{-\infty}^{\infty} U^2\, dx}{\displaystyle\int_{-\infty}^{\infty} U^2\, dx} = \hbar \mathbf{k}_x$$

With these limits of integration, both the numerator and denominator are infinite. For problems of this type, one integrates between the finite limits $-L/2$ and $+L/2$ and, in the final result, then assumes that L approaches infinity.

This result implies that $(E, \mathbf{k})$ diagrams such as shown in Fig. 3-5 can be considered plots of electron energy vs. momentum, with a scaling factor $\hbar$.

The direct and indirect semiconductors are identified in Appendix III. In a direct semiconductor such as GaAs, an electron in the conduction band can fall to an empty state in the valence band, giving off the energy difference E_g as a photon of light. On the other hand, an electron in the conduction band minimum of an indirect semiconductor such as Si cannot fall directly to the valence band maximum but must undergo a momentum change as well as changing its energy. For example, it may go through some defect state (E_t) within the band gap. We shall discuss such defect states in Sections 4.2.1 and 4.3.2. In an indirect transition which involves a change in $\mathbf{k}$, part of the energy is generally given up as heat to the lattice rather than as an emitted photon. This difference between direct and indirect band structures is very important for deciding which semiconductors can be used in devices requiring light output. For example, semiconductor light emitters and lasers (Chapter 8) generally must be made of materials capable of direct band-to-band transitions or of indirect materials with vertical transitions between defect states.

Band diagrams such as those shown in Fig. 3-5 are cumbersome to draw in analyzing devices, and do not provide a view of the variation of electron energy with distance in the sample. Therefore, in most discussions we shall use simple band pictures such as those shown in Fig. 3-4, remembering that electron transitions across the band gap may be direct or indirect.

3.1.5 Variation of Energy Bands with Alloy Composition

As III–V ternary and quaternary alloys are varied over their composition ranges (see Sections 1.2.4 and 1.4.1), their band structures change. For example, Fig. 3-6 illustrates the band structure of GaAs and AlAs, and the way in which the bands change with composition x in the ternary compound $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$. The binary compound GaAs is a direct material, with a band gap of 1.43 eV at room temperature. For reference, we call the direct ($\mathbf{k} = 0$) conduction band minimum Γ. There are also two higher-lying indirect minima in the GaAs conduction band, but these are sufficiently far above Γ that few electrons reside there (we discuss an important exception in Chapter 10 in which high-field excitation of electrons into the indirect minima leads to the Gunn effect). We call the lowest-lying GaAs indirect minimum L and the other X. In AlAs the direct Γ minimum is much higher than the indirect X minimum, and this material is therefore indirect with a band gap of 2.16 eV at room temperature.

In the ternary alloy $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$ all of these conduction band minima move up relative to the valence band as the composition x varies from 0 (GaAs) to 1 (AlAs). However, the indirect minimum X moves up less than the others, and for compositions above about 38 percent Al this indirect minimum becomes the lowest-lying conduction band. Therefore, the ternary alloy AlGaAs is a direct semiconductor for Al compositions on the column III sublattice up to about 38 percent, and is an indirect semiconductor for higher Al mole fractions. The band gap energy E_g is shown in color on Fig. 3-6c.

The variation of energy bands for the ternary alloy $\mathrm{GaAs}_{1-x}\mathrm{P}_x$ is generally similar to that of AlGaAs shown in Fig. 3-6. GaAsP is a direct semiconductor from GaAs to about $\mathrm{GaAs}_{.55}\mathrm{P}_{.45}$ and is indirect from this composition to GaP (see Fig. 8-11). This material is often used in visible LEDs.

Since light emission is most efficient for direct materials, in which electrons can drop from the conduction band to the valence band without changing $\mathbf{k}$ (and therefore momentum), LEDs in GaAsP are generally made in material grown with a composition less than $x = 0.45$. For example, most red LEDs in this material are made at about $x = 0.4$, where the Γ minimum is still the lowest-lying conduction band edge, and where the photon resulting from a direct transition from this band to the valence band is in the red portion of the spectrum (about 1.9 eV). The use of impurities to enhance radiative recombination in indirect material will be discussed in Section 8.2.

3.2
CHARGE CARRIERS IN SEMICONDUCTORS

The mechanism of current conduction is relatively easy to visualize in the case of a metal; the metal atoms are imbedded in a "sea" of relatively free electrons, and these electrons can move as a group under the influence of an electric field. This free electron view is oversimplified, but many important conduction properties of metals can be derived from just such a model. However, we cannot account for all of the electrical properties of semiconductors

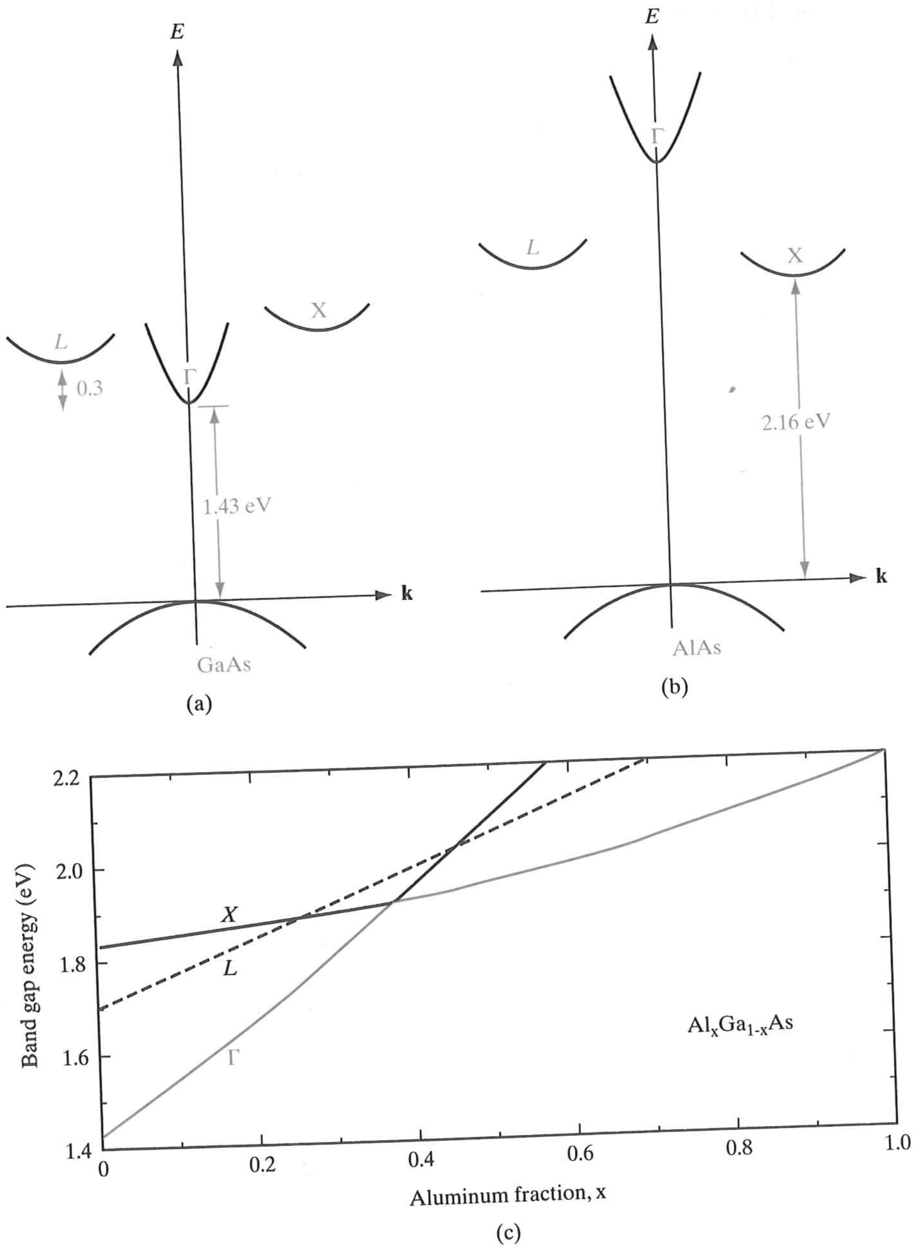


Figure 3–6
Variation of direct and indirect conduction bands in AlGaAs as a function of composition: (a) the $(E,\mathbf{k})$ diagram for GaAs, showing three minima in the conduction band; (b) AlAs band diagram; (c) positions of the three conduction band minima in $Al_xGa_{1-x}As$ as x varies over the range of compositions from GaAs ($x = 0$) to AlAs ($x = 1$). The smallest band gap, E_g (shown in color), follows the direct Γ band to $x = 0.38$, and then follows the indirect X band.

in this way. Since the semiconductor has a filled valence band and an empty conduction band at 0 K, we must consider the increase in conduction band electrons by thermal excitations across the band gap as the temperature is raised. In addition, after electrons are excited to the conduction band, the empty states left in the valence band can contribute to the conduction process. The introduction of impurities has an important effect on the energy band structure and on the availability of charge carriers. Thus there is considerable flexibility in controlling the electrical properties of semiconductors.

3.2.1 Electrons and Holes

As the temperature of a semiconductor is raised from 0 K, some electrons in the valence band receive enough thermal energy to be excited across the band gap to the conduction band. The result is a material with some electrons in an otherwise empty conduction band and some unoccupied states in an otherwise filled valence band (Fig. 3–7).[2] For convenience, an empty state in the valence band is referred to as a *hole*. If the conduction band electron and the hole are created by the excitation of a valence band electron to the conduction band, they are called an *electron–hole pair* (abbreviated EHP).

After excitation to the conduction band, an electron is surrounded by a large number of unoccupied energy states. For example, the equilibrium number of electron–hole pairs in pure Si at room temperature is only about 10^{10} EHP/cm^3, compared to the Si atom density of 5×10^{22} atoms/cm^3. Thus the few electrons in the conduction band are free to move about via the many available empty states.

The corresponding problem of charge transport in the valence band is somewhat more complicated. However, it is possible to show that the effects of current in a valence band containing holes can be accounted for by simply keeping track of the holes themselves.

In a filled band, all available energy states are occupied. For every electron moving with a given velocity, there is an equal and opposite electron motion elsewhere in the band. If we apply an electric field, the net current is zero

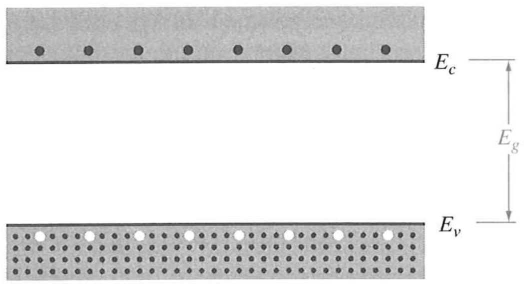


Figure 3–7
Electron–hole pairs in a semiconductor.

---

[2] In Fig. 3–7 and in subsequent discussions, we refer to the bottom of the conduction band as E_c and the top of the valence band as E_v.

because for every electron j moving with velocity v_j there is a corresponding electron j' with velocity $-\mathrm{v}_j$. Figure 3-8 illustrates this effect in terms of the electron energy vs. wave vector plot for the valence band. Since **k** is proportional to electron momentum, it is clear the two electrons have oppositely directed velocities. With N electrons/cm^3 in the band we express the current density using a sum over all of the electron velocities, and including the charge $-q$ on each electron. In a unit volume,

$$J = (-q)\sum_{i}^{N} \mathrm{v}_i = 0 \quad (\textit{filled band}) \tag{3-2a}$$

Now if we create a hole by removing the jth electron, the net current density in the valence band involves the sum over all velocities, minus the contribution of the electron we have removed:

$$J = (-q)\sum_{i}^{N} \mathrm{v}_i - (-q)\mathrm{v}_j \quad (j\text{th } \textit{electron missing}) \tag{3-2b}$$

But the first term is zero, from Eq. (3-2a). Thus the net current is $+q\mathrm{v}_j$. In other words, the current contribution of the hole is equivalent to that of a positively charged particle with velocity v_j, that of the missing electron. Of course, the charge transport is actually due to the motion of the new uncompensated electron (j'). Its current contribution $(-q)\,(-\mathrm{v}_j)$ is equivalent to that of a positively charged particle with velocity $+\mathrm{v}_j$. For simplicity, it is customary to treat empty states in the valence band as charge carriers with positive charge and positive mass.

A simple analogy may help in understanding the behavior of holes. If we have two bottles, one completely filled with water and one completely empty, we can ask ourselves "Will there be any net transport of water when we tilt the bottles?" The answer is "no". In the case of the empty bottle, the answer is obvious. In the case of the completely full bottle also, there cannot be any net motion of water because there is no empty space for water to move into. Similarly, an empty conduction band completely devoid of electrons or a valence band completely full of electrons cannot give rise to a net motion of electrons, and thus to current conduction.

Next, we imagine transferring some water droplets from the full bottle into the empty bottle, leaving behind some air bubbles, and ask ourselves the same question. Now when we tilt the bottles there will be net transport of water: the water droplets will roll downhill in one bottle and the air bubbles will move uphill in the other. Similarly, a few electrons in an otherwise empty conduction band move opposite to an electric field, while holes in an otherwise filled valence band move in the direction of the field. The bubble analogy is imperfect, but it may provide a physical feel for why the charge and mass of a hole have opposite signs from those of an electron.

In all the following discussions we shall concentrate on the electrons in the conduction band and on the holes in the valence band. We can account for the current flow in a semiconductor by the motion of these two types of charge carriers. We draw valence and conduction bands on an electron energy scale E, as in Fig. 3-8. However, we should remember that in the valence band, hole energy increases oppositely to electron energy, because the two carriers have opposite charge. Thus hole energy increases downward in Fig. 3-8 and holes, seeking the lowest energy state available, are generally found at the *top* of the valence band. In contrast, conduction band electrons are found at the bottom of the conduction band.

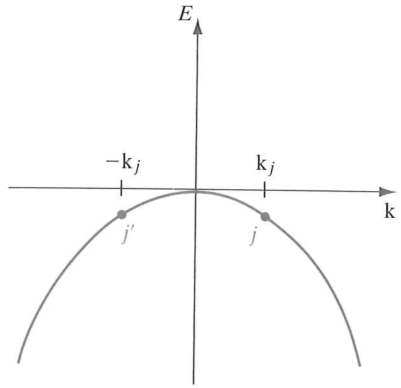
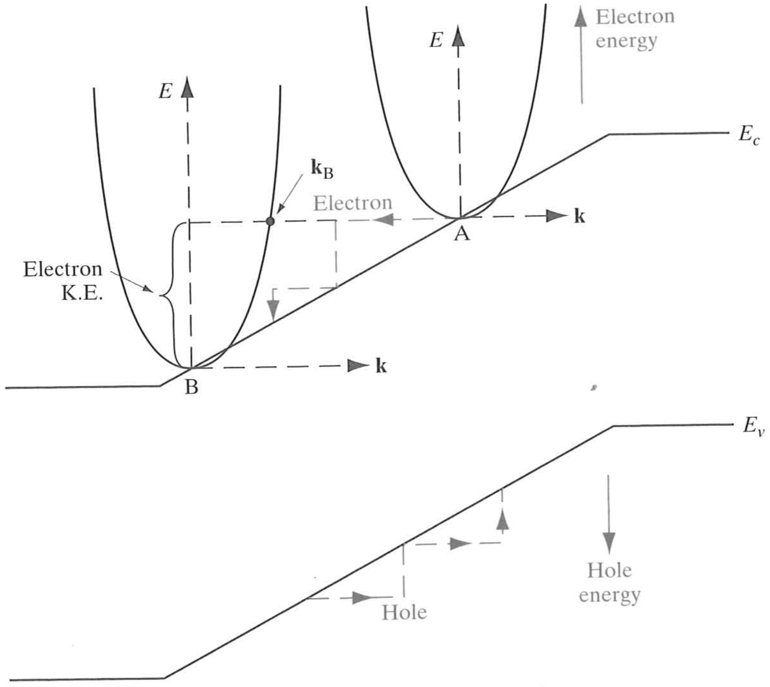


Figure 3–8
A valence band with all states filled, including states j and j', marked for discussion. The jth electron with wave vector k_j is matched by an electron at j' with the opposite wave vector $-\mathrm{k}_j$. There is no net current in the band unless an electron is removed. For example, if the jth electron is removed, the motion of the electron at j' is no longer compensated.

It would be instructive to compare the $(E, \mathbf{k})$ band diagrams with the "simplified" band diagrams that are used for routine device analysis (Fig. 3-9). As discussed in Examples 3-1 and 3-2, an $(E, \mathbf{k})$ diagram is a plot of the total electron energy (potential plus kinetic) as a function of the crystal-direction-dependent electron wave vector (which is proportional to the momentum and therefore the velocity) at some point in space. Hence, the bottom of the conduction band corresponds to zero electron velocity or kinetic energy, and simply gives us the potential energy at that point in space. For holes, the top of the valence band corresponds to zero kinetic energy. For simplified band diagrams, we plot the edges of the conduction and valence bands (i.e., the potential energy) as a function of position in the device. Energies higher in the band correspond to additional kinetic energy of the electron. Also, the fact that the band edge corresponds to the electron potential energy tells us that the variation of the band edge in space is related to the electric field at different points in the semiconductor. We will show this relationship explicitly in Section 4.4.2.

In Fig. 3-9, an electron at location A sees an electric field given by the slope of the band edge (potential energy), and gains kinetic energy (at the expense of potential energy) by moving to point B. Correspondingly, in the $(E, \mathbf{k})$ diagram, the electron starts at $k = 0$, but moves to a nonzero wave vector $\mathbf{k}_{\mathrm{B}}$.

Figure 3–9 Superimposition of the $(E,\mathbf{k})$ band structure on the *E*-versus-position simplified band diagram for a semiconductor in an electric field. Electron energies increase going up, while hole energies increase going down. Similarly, electron and hole wave vectors point in opposite directions and these charge carriers move opposite to each other, as shown.

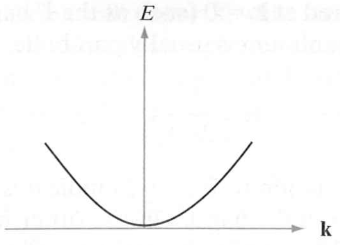


The electron then loses kinetic energy to heat by scattering mechanisms (discussed in Section 3.4.3) and returns to the bottom of the band at B. The slopes of the (E, x) band edges at different points in space reflect the local electric fields at those points. In practice, the electron may lose its kinetic energy in stages by a series of scattering events, as shown by the colored dashed lines.

3.2.2 Effective Mass

The electrons in a crystal are not completely free, but instead interact with the periodic potential of the lattice. As a result, their "wave-particle" motion cannot be expected to be the same as for electrons in free space. Thus, in applying the usual equations of electrodynamics to charge carriers in a solid, we must use altered values of particle mass. In doing so, we account for most of the influences of the lattice, so that the electrons and holes can be treated as "almost free" carriers in most computations. The calculation of effective mass must take into account the shape of the energy bands in three-dimensional **k**-space, taking appropriate averages over the various energy bands.

EXAMPLE 3–2 Find the $(E, \mathbf{k})$ relationship for a free electron and relate it to the electron mass.



SOLUTION From Example 3-1, the electron momentum is $\mathrm{p} = m\mathrm{v} = \hbar\mathbf{k}$. Then

$$E = \frac{1}{2}m\mathrm{v}^2 = \frac{1}{2}\frac{\mathrm{p}^2}{m} = \frac{\hbar^2}{2m}\mathbf{k}^2$$

Thus the electron energy is parabolic with wave vector **k**. The electron mass is inversely related to the curvature (second derivative) of the $(E, \mathbf{k})$ relationship, since

$$\frac{d^2E}{d\mathbf{k}^2} = \frac{\hbar^2}{m}$$

Although electrons in solids are not free, most energy bands are close to parabolic at their minima (for conduction bands) or maxima (for valence bands). We can also approximate effective mass near those band extrema from the curvature of the band.

The effective mass of an electron in a band with a given $(E, \mathbf{k})$ relationship is found in Example 3-2 to be

$$m* = \frac{\hbar^2}{d^2E/d\mathbf{k}^2} \tag{3-3}$$

Thus the curvature of the band determines the electron effective mass. For example, in Fig. 3-6a it is clear that the electron effective mass in GaAs is much smaller in the direct Γ conduction band (strong curvature) than in the *L* or *X* minima (weaker curvature, smaller value in the denominator of the $m*$ expression).

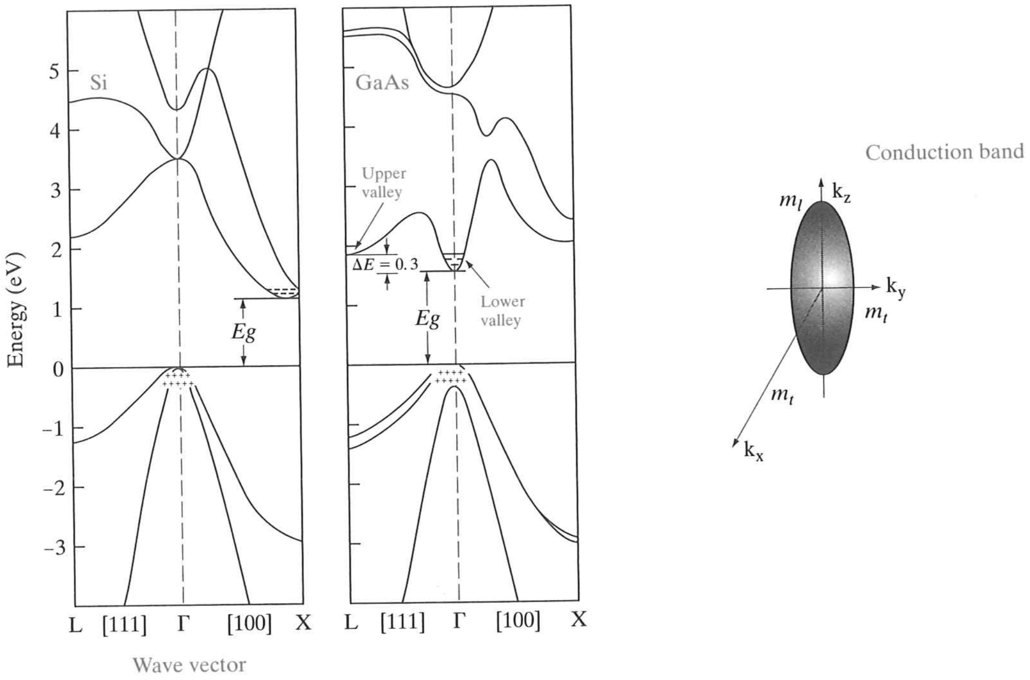
A particularly interesting feature of Figs. 3-5 and 3-6 is that the curvature of $d^2E/d\mathbf{k}^2$ is positive at the conduction band minima, but is negative at the valence band maxima. Thus, the electrons near the top of the valence band have *negative effective mass,* according to Eq. (3-3). Valence band electrons with negative charge and negative mass move in an electric field in the same direction as holes with positive charge and positive mass. As discussed in Section 3.2.1, we can fully account for charge transport in the valence band by considering hole motion.

For a band centered at $\mathbf{k} = 0$ (such as the Γ band in GaAs), the $(E, \mathbf{k})$ relationship near the minimum is usually parabolic:

$$E = \frac{\hbar^2}{2m^*}\mathbf{k}^2 + E_c \qquad (3\text{-}4)$$

Comparing this relation to Eq. (3-3) indicates that the effective mass m^* is constant in a parabolic band. On the other hand, many conduction bands have complex $(E, \mathbf{k})$ relationships that depend on the direction of electron transport with respect to the principal crystal directions. In this case, the effective mass is a tensor quantity. However, we can use appropriate averages over such bands in most calculations.

Figure 3–10a shows the band structures for Si and GaAs viewed along two major directions. While the shape is parabolic near the band edges (as indicated in Figure 3-5 and Example 3-2), there are significant non-parabolicities at higher energies. The energies are plotted along the high symmetry [111] and [100] directions in the crystal. The $\mathbf{k} = 0$ point is denoted as Γ. When we go along the [100] direction, we reach a valley near X, while we reach the L valley along the [111] direction. (Since the energies are plotted along different directions, the curves do not look symmetric.) The valence band maximum in most semiconductors is at the Γ point. It has three branches: the *heavy hole* band with the smallest curvature, a *light hole band* with a larger curvature, and a *split-off band* at a different energy. We notice that for GaAs the conduction band minimum and the valence band maximum are both at $\mathbf{k} = 0$; therefore it is direct band gap. Silicon, on the other hand, has 6 equivalent conduction minima at X along the 6 equivalent ⟨100⟩ directions; therefore, it is indirect.



(a)

(b)

Figure 3–10
Realistic band structures in semiconductors: (a) Conduction and valence bands in Si and GaAs along [111] and [100]; (b) ellipsoidal constant energy surface for Si, near the 6 conduction band minima along the X directions. (From Chelikowsky and Cohen, Phys. Rev. B14, 556, 1976).

Figure 3–10b shows the constant energy surface for electrons in one of the six conduction bands for Si. The way to relate these surfaces to the band structures shown in Fig. 3–10a is to consider a certain value of energy, and determine all the $\mathbf{k}$ vectors in 3 dimensions for which we get this energy. We find that for Si we have 6 cigar-shaped ellipsoidal equi-energy surfaces near the conduction band minima along the six equivalent X-directions, with a longitudinal effective mass, m_l, along the major axis, and two transverse effective masses, m_t, along the minor axes. For GaAs, the conduction band is more or less spherical for low energies. On the other hand, we have warped spherical surfaces in the valence band. The importance of these surfaces will be clear in Sections 3.3.2 and 3.4.1 when we consider different types of effective masses in semiconductors.

In any calculation involving the mass of the charge carriers, we must use effective mass values for the particular material involved. In all subsequent discussions, the electron effective mass is denoted by m_n^* and the hole effective mass by m_p^*. The n subscript indicates the electron as a negative charge carrier, and the p subscript indicates the hole as a positive charge carrier.

There is nothing mysterious about the concept of an "effective" mass, m_n^*, and about the fact that it is different in different semiconductors. Indeed, the "true" mass of an electron, m, is the same in Si, Ge, or GaAs—it is the same as for a free electron in vacuum. To understand why the effective mass is different from the true mass, consider Newton's second law, which states that the time rate of change of momentum is the force.

$$dp/dt = d(m\mathrm{v})/dt = \text{Force} \qquad (3\text{-}5a)$$

An electron in a crystal experiences a total force $F_{int} + F_{ext}$, where F_{int} is the collection of internal periodic crystal forces, and F_{ext} is the externally applied force. It is inefficient to solve this complicated problem involving the periodic crystal potential (which is obviously different in different semiconductors) every time we try to solve a semiconductor device problem. It is better to solve the complicated problem of carrier motion in the periodic crystal potential just once, and encapsulate that information in what is called the bandstructure, $(E, \mathbf{k})$, whose curvature gives us the effective mass, m_n^*. The electron then responds to external forces with this new m_n^*. Newton's law is then written as:

$$d(m_n^*\mathrm{v})/dt = F_{ext} \qquad (3\text{-}5b)$$

This is clearly an enormous simplification compared to the more detailed problem. Obviously, the periodic crystal forces depend on the details of a specific semiconductor; therefore, the effective mass is different in different materials.

Once we determine the band curvature effective mass components from the orientation-dependent (E, $\mathbf{k}$), we have to combine them appropriately for different types of calculations. We shall see in Section 3.3.2 that when we are interested in determining the numbers of carriers in the bands, we have to use a "density-of-states" effective mass by taking the geometric mean of the band curvature effective masses, and the number of equivalent band extrema. On the other hand we will find in Section 3.4.1 that in problems involving the motion of carriers, one must take the harmonic mean of the band curvature effective masses to get the "conductivity" effective mass.

3.2.3 Intrinsic Material

A perfect semiconductor crystal with no impurities or lattice defects is called an *intrinsic* semiconductor. In such material there are no charge carriers at 0 K, since the valence band is filled with electrons and the conduction band is empty. At higher temperatures electron–hole pairs are generated as valence band electrons are excited thermally across the band gap to the conduction band. These EHPs are the only charge carriers in intrinsic material.

The generation of EHPs can be visualized in a qualitative way by considering the breaking of covalent bonds in the crystal lattice (Fig. 3–11). If one of the Si valence electrons is broken away from its position in the bonding structure such that it becomes free to move about in the lattice, a conduction electron is created and a broken bond (hole) is left behind. The energy required to break the bond is the band gap energy E_g. This model helps in visualizing the physical mechanism of EHP creation, but the energy band model is more productive for purposes of quantitative calculation. One important difficulty in the "broken bond" model is that the free electron and the hole seem deceptively localized in the lattice. Actually, the positions of the free electron and the hole are spread out over several lattice spacings and should be considered quantum mechanically by probability distributions (see Section 2.4).

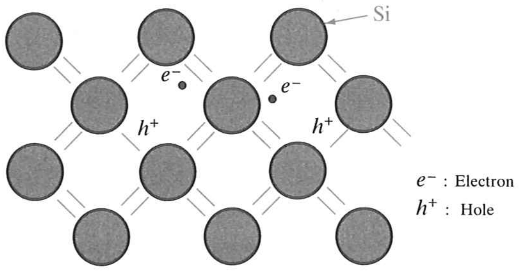


Figure 3–11 Electron–hole pairs in the covalent bonding model of the Si crystal.

Since the electrons and holes are created in pairs, the conduction band electron concentration n (electrons per cm^3) is equal to the concentration of holes in the valence band p (holes per cm^3). Each of these intrinsic carrier concentrations is commonly referred to as n_i. Thus *for intrinsic material*

$$n = p = n_i \tag{3–6}$$

At a given temperature there is a certain concentration of electron–hole pairs n_i. Obviously, if a steady state carrier concentration is maintained, there must be *recombination* of EHPs at the same rate at which they are generated. Recombination occurs when an electron in the conduction band makes a transition (direct or indirect) to an empty state (hole) in the valence band, thus annihilating the pair. If we denote the generation rate of EHPs as g_i (EHP/cm^3-s) and the recombination rate as r_i, equilibrium requires that

$$r_i = g_i \tag{3–7a}$$

Each of these rates is temperature dependent. For example, $g_i(T)$ increases when the temperature is raised, and a new carrier concentration n_i is established such that the higher recombination rate $r_i(T)$ just balances generation. At any temperature, we can predict that the rate of recombination of electrons and holes r_i is proportional to the equilibrium concentration of electrons n_0 and the concentration of holes p_0:

$$r_i = \alpha_r n_0 p_0 = \alpha_r n_i^2 = g_i \tag{3–7b}$$

The factor α_r is a constant of proportionality which depends on the particular mechanism by which recombination takes place. We shall discuss the calculation of n_i as a function of temperature in Section 3.3.3; recombination processes will be discussed in Chapter 4.

3.2.4 Extrinsic Material

In addition to the intrinsic carriers generated thermally, it is possible to create carriers in semiconductors by purposely introducing impurities into the crystal. This process, called *doping*, is the most common technique for varying the conductivity of semiconductors. By doping, a crystal can be altered so that it has a predominance of either electrons or holes. Thus there are two types of doped semiconductors, n-type (mostly electrons) and p-type (mostly holes). When a crystal is doped such that the equilibrium carrier concentrations n_0 and p_0 are different from the intrinsic carrier concentration n_i, the material is said to be *extrinsic*.

When impurities or lattice defects are introduced into an otherwise perfect crystal, additional levels are created in the energy band structure, usually within the band gap. For example, an impurity from column V of the periodic table (P, As, and Sb) introduces an energy level very near the conduction band in Ge or Si. This level is filled with electrons at 0 K, and very little thermal energy is required to excite these electrons to the conduction band (Fig. 3–12a). Thus at about 50–100 K virtually all of the electrons in the

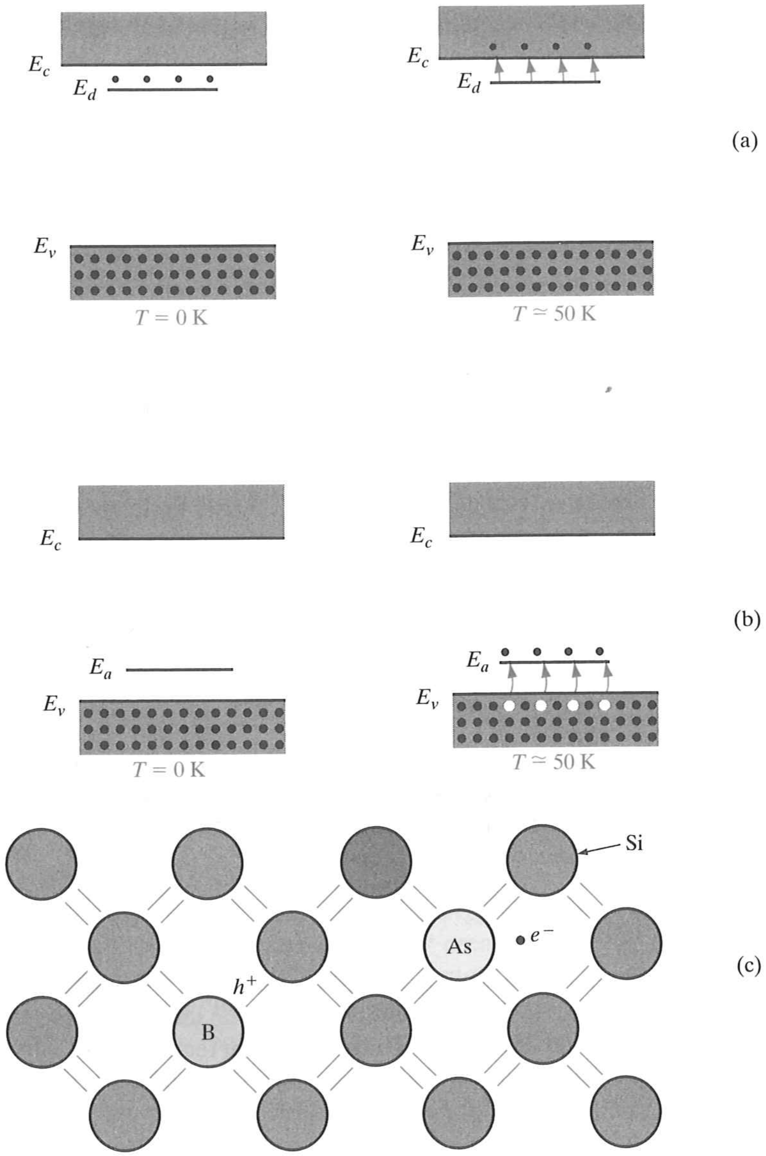


Figure 3–12 Energy band model and chemical bond model of dopants in semiconductors: (a) donation of electrons from donor level to conduction band; (b) acceptance of valence band electrons by an acceptor level, and the resulting creation of holes; (c) donor and acceptor atoms in the covalent bonding model of a Si crystal.

impurity level are "donated" to the conduction band. Such an impurity level is called a *donor* level, and the column V impurities in Ge or Si are called donor impurities. From Fig. 3–12a we note that the material doped with donor impurities can have a considerable concentration of electrons in the conduction band, even when the temperature is too low for the intrinsic EHP concentration to be appreciable. Thus semiconductors doped with a significant number of donor atoms will have $n_0 \gg (n_i, p_0)$ at room temperature. This is n-type material.

Atoms from column III (B, Al, Ga, and In) introduce impurity levels in Ge or Si near the valence band. These levels are empty of electrons at 0 K (Fig. 3–12b). At low temperatures, enough thermal energy is available to excite electrons from the valence band into the impurity level, leaving behind holes in the valence band. Since this type of impurity level "accepts" electrons from the valence band, it is called an *acceptor* level, and the column III impurities are acceptor impurities in Ge and Si. As Fig. 3–12b indicates, doping with acceptor impurities can create a semiconductor with a hole concentration p_0 much greater than the conduction band electron concentration n_0 (this type is p-type material).

In the covalent bonding model, donor and acceptor atoms can be visualized as shown in Fig. 3–12c. An As atom (column V) in the Si lattice has the four necessary valence electrons to complete the covalent bonds with the neighboring Si atoms, plus one extra electron. This fifth electron does not fit into the bonding structure of the lattice and is therefore loosely bound to the As atom. A small amount of thermal energy enables this extra electron to overcome its coulombic binding to the impurity atom and be donated to the lattice as a whole. Thus it is free to participate in current conduction. This process is a qualitative model of the excitation of electrons out of a donor level and into the conduction band (Fig. 3–12a). Similarly, the column III impurity B has only three valence electrons to contribute to the covalent bonding (Fig. 3–12c), thereby leaving one bond incomplete. With a small amount of thermal energy, this incomplete bond can be transferred to other atoms as the bonding electrons exchange positions. Again, the idea of an electron "hopping" from an adjacent bond into the incomplete bond at the B site provides some physical insight into the behavior of an acceptor, but the model of Fig. 3–12b is preferable for most discussions.

We can calculate rather simply the approximate energy required to excite the fifth electron of a donor atom into the conduction band (the donor *binding energy*). Let us assume for rough calculations that the As atom of Fig. 3–12c has its four covalent bonding electrons rather tightly bound and the fifth "extra" electron loosely bound to the atom. We can approximate this situation by using the Bohr model results, considering the loosely bound electron as ranging about the tightly bound "core" electrons in a hydrogen-like orbit. From Eq. (2–15) the magnitude of the ground-state energy ($\mathbf{n} = 1$) of such an electron is

$$E = \frac{mq^4}{2K^2\hbar^2} \tag{3-8}$$

The value of K must be modified from the free-space value $4\pi\epsilon_0$ used in the hydrogen atom problem to

$$K = 4\pi\epsilon_0\,\epsilon_r \tag{3-9}$$

where ϵ_r is the relative dielectric constant of the semiconductor material. In addition, we must use the conductivity effective mass m_n^* typical of the semiconductor, discussed in more detail in Section 3.4.1.

EXAMPLE 3–3 It was mentioned in Section 3.2 that the covalent bonding model gives a false impression of the localization of carriers. As an illustration, calculate the radius of the electron orbit around the donor in Fig. 3–12c, assuming a ground state hydrogen-like orbit in Si. Compare with the Si lattice constant. Use $m_n^* = 0.26m_0$ for Si.

SOLUTION From Eq. (2–10) with $n = 1$ and using $\epsilon_r = 11.8$ for Si,

$$r = \frac{4\pi\,\epsilon_r\epsilon_0\hbar^2}{m_n^* q^2} = \frac{11.8(8.85 \times 10^{-12})(6.63 \times 10^{-34})^2}{\pi(0.26)(9.11 \times 10^{-31})(1.6 \times 10^{-19})^2}$$

$$r = 2.41 \times 10^{-9} m = 24.1\ \text{Å}.$$

Note that this is more than four lattice spacings $a = 5.43$ Å.

Generally, the column V donor levels lie approximately 0.01 eV below the conduction band in Ge, and the column III acceptor levels lie about 0.01 eV above the valence band. In Si the usual donor and acceptor levels lie about 0.03–0.06 eV from a band edge.

In III–V compounds, column VI impurities occupying column V sites serve as donors. For example, S, Se, and Te are donors in GaAs, since they substitute for As and provide an extra electron compared with the As atom. Similarly, impurities from column II (Be, Zn, Cd) substitute for column III atoms to form acceptors in the III–V compounds. A more ambiguous case arises when a III–V material is doped with Si or Ge, from column IV. These impurities are called *amphoteric,* meaning that Si or Ge can serve as donors or acceptors depending on whether they reside on the column III or column V sublattice of the crystal. In GaAs it is common for Si impurities to occupy Ga sites. Since the Si has an extra electron compared with the Ga it replaces, it serves as a donor. However, an excess of As vacancies arising during growth or processing of the GaAs can cause Si impurities to occupy As sites, where they serve as acceptors.

The importance of doping will become obvious when we discuss electronic devices made from junctions between p-type and n-type semiconductor material. The extent to which doping controls the electronic properties of semiconductors can be illustrated here by considering changes in the sample resistance which occur with doping. In Si, for example, the intrinsic carrier concentration n_i is about 10^{10} cm^{-3} at room temperature. If we dope Si with 10^{15} As atoms/cm^3, the conduction electron concentration changes by five orders of magnitude. The resistivity of Si changes from about 2×10^5 Ω-cm to 5 Ω-cm with this doping.

When a semiconductor is doped n-type or p-type, one type of carrier dominates. In the example given above, the conduction band electrons outnumber the holes in the valence band by many orders of magnitude. We refer to the small number of holes in n-type material as *minority carriers* and the relatively large number of conduction band electrons as *majority carriers.* Similarly, electrons are the minority carriers in p-type material, and holes are the majority carriers.

3.2.5 Electrons and Holes in Quantum Wells

We have discussed single-valued (*discrete*) energy levels in the band gap arising from doping, and a *continuum* of allowed states in the valence and conduction bands. A third possibility is the formation of discrete levels for electrons and holes as a result of quantum-mechanical confinement.

One of the most useful applications of MBE or OMVPE growth of multi-layer compound semiconductors, as described in Section 1.4, is the fact that a continuous single crystal can be grown in which adjacent layers have different band gaps. For example, Fig. 3–13 shows the spatial variation in conduction and valence bands for a multilayer structure in which a very thin layer of GaAs is sandwiched between two layers of AlGaAs, which has a wider band gap that the GaAs. We will discuss the details of such *heterojunctions* (junctions between dissimilar materials) in Section 5.8. It is interesting to point out here, however, that a consequence of confining electrons and holes in a very thin layer is that these particles behave according to the *particle in a potential well* problem, with quantum states calculated in Section 2.4.3. Therefore,

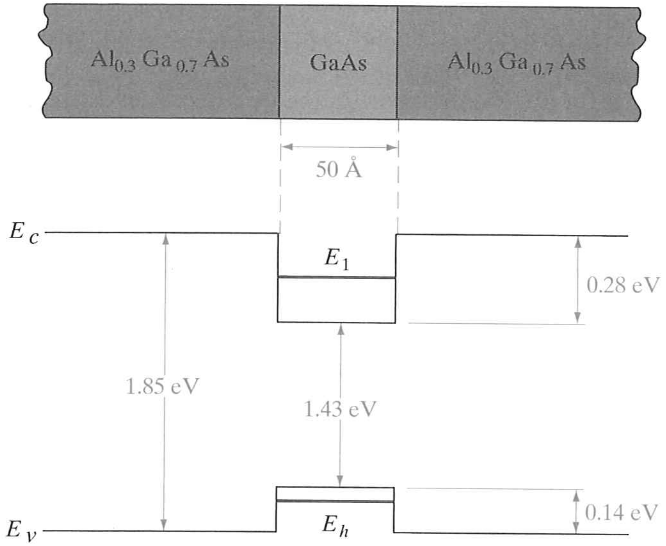


Figure 3–13
Energy band discontinuities for a thin layer of GaAs sandwiched between layers of wider band gap AlGaAs. In this case, the GaAs region is so thin that quantum states are formed in the valence and conduction bands. Electrons in the GaAs conduction band reside on "particle in a potential well" states such as E_1 shown here, rather than in the usual conduction band states. Holes in the quantum well occupy similar discrete states, such as E_h.

instead of having the continuum of states normally available in the conduction band, the conduction band electrons in the narrow-gap material are confined to discrete quantum states as described by Eq. (2–33), modified for effective mass and finite barrier height. Similarly, the states in the valence band available for holes are restricted to discrete levels in the quantum well. This is one of the clearest demonstrations of the quantum mechanical results discussed in Chapter 2. From a practical device point of view, the formation of discrete quantum states in the GaAs layer of Fig. 3–13 changes the energy at which photons can be emitted. An electron on one of the discrete conduction band states (E_1 in Fig. 3–13) can make a transition to an empty discrete valence band state in the GaAs quantum well (such as E_h), giving off a photon of energy $E_g + E_1 + E_h$, greater than the GaAs band gap. Semiconductor lasers have been made in which such a quantum well is used to raise the energy of the transition from the infrared, typical of GaAs, to the red portion of the spectrum. We will see other examples of quantum wells in semiconductor devices in later chapters.

3.3 CARRIER CONCENTRATIONS

In calculating semiconductor electrical properties and analyzing device behavior, it is often necessary to know the number of charge carriers per cm^3 in the material. The majority carrier concentration is usually obvious in heavily doped material, since one majority carrier is obtained for each impurity atom (for the standard doping impurities). The concentration of minority carriers is not obvious, however, nor is the temperature dependence of the carrier concentrations.

To obtain equations for the carrier concentrations we must investigate the distribution of carriers over the available energy states. This type of distribution is not difficult to calculate, but the derivation requires some background in statistical methods. Since we are primarily concerned here with the application of these results to semiconductor materials and devices, we shall accept the distribution function as given.

3.3.1 The Fermi Level

Electrons in solids obey *Fermi–Dirac* statistics.[3] In the development of this type of statistics, one must consider the indistinguishability of the electrons,

---

[3] Examples of other types of statistics are *Maxwell–Boltzmann* for classical particles (e.g., gas) and *Bose–Einstein* for photons. For two discrete energy levels, E_2 and E_1 (with $E_2 > E_1$), classical gas atoms follow a Boltzmann distribution; the number n_2 of atoms in state E_2 is related to the number n_1 in E_1 at thermal equilibrium by

$$\frac{n_2}{n_1} = \frac{N_2 e^{-E_2/kT}}{N_1 e^{-E_1/kT}} = \frac{N_2}{N_1} e^{-(E_2 - E_1)/kT}$$

assuming the two levels have N_2 and N_1 number of states, respectively. The exponential term $\exp(-\Delta E/kT)$ is commonly called the *Boltzmann factor.* It appears also in the denominator of the Fermi–Dirac distribution function. We shall return to the Boltzmann distribution in Chapter 8 in discussions of the properties of lasers.

their wave nature, and the Pauli exclusion principle. The rather simple result of these statistical arguments is that the distribution of electrons over a range of allowed energy levels at thermal equilibrium is

$$\boxed{f(E) = \frac{1}{1 + e^{(E-E_F)/kT}}} \qquad (3\text{–}10)$$

where k is Boltzmann's constant ($k = 8.62 \times 10^{-5}$ eV/K $= 1.38 \times 10^{-23}$ J/K). The function $f(E)$, the *Fermi–Dirac distribution function,* gives the probability that an available energy state at E will be occupied by an electron at absolute temperature T. The quantity E_F is called the *Fermi level,* and it represents an important quantity in the analysis of semiconductor behavior. We notice that, for an energy E equal to the Fermi level energy E_F, the occupation probability is

$$f(E_F) = [1 + e^{(E_F - E_F)/kT}]^{-1} = \frac{1}{1+1} = \frac{1}{2} \qquad (3\text{–}11)$$

Thus an energy state at the Fermi level has a probability of $^1/_2$ of being occupied by an electron.

A closer examination of $f(E)$ indicates that at 0 K the distribution takes the simple rectangular form shown in Fig. 3–14. With $T = 0$ in the denominator of the exponent, $f(E)$ is $1/(1 + 0) = 1$ when the exponent is negative ($E < E_F$), and is $1/(1 + \infty) = 0$ when the exponent is positive ($E > E_F$). This rectangular distribution implies that at 0 K every available energy state up to E_F is filled with electrons, and all states above E_F are empty.

At temperatures higher than 0 K, some probability exists for states above the Fermi level to be filled. For example, at $T = T_1$ in Fig. 3–14 there is some probability f(E) that states above E_F are filled, and there is a corresponding probability $[1 - f(E)]$ that states below E_F are empty. The Fermi function is symmetrical about E_F for all temperatures; that is, the probability $f(E_F + \Delta E)$ that a state ΔE above E_F is filled is the same as the probability $[1 - f(E_F - \Delta E)]$ that a state ΔE below E_F is empty. The symmetry of the

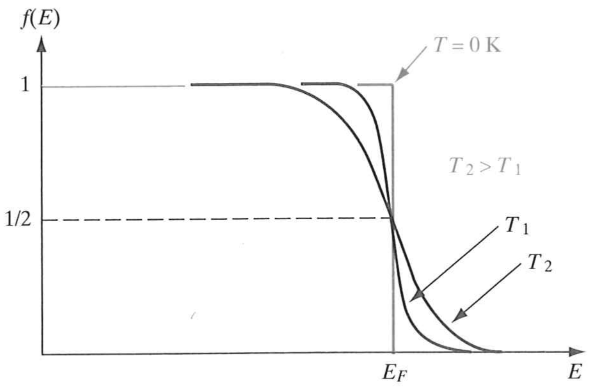
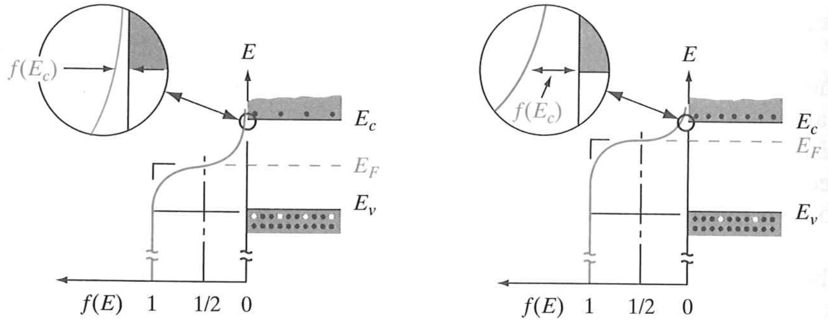


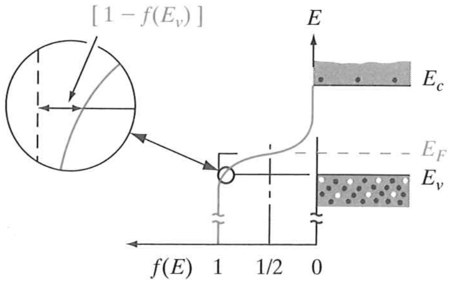
Figure 3–14 The Fermi–Dirac distribution function.

distribution of empty and filled states about E_F makes the Fermi level a natural reference point in calculations of electron and hole concentrations in semiconductors.

In applying the Fermi–Dirac distribution to semiconductors, we must recall that $f(E)$ is the probability of occupancy of an *available* state at E. Thus if there is no available state at E (e.g., in the band gap of a semiconductor), there is no possibility of finding an electron there. We can best visualize the relation between $f(E)$ and the band structure by turning the $f(E)$ vs. E diagram on its side so that the E scale corresponds to the energies of the band diagram (Fig. 3–15). For intrinsic material we know that the concentration of holes in the valence band is equal to the concentration of electrons in the conduction band. Therefore, the Fermi level E_F must lie at the middle of the band gap in intrinsic material.[4] Since $f(E)$ is symmetrical about E_F, the electron probability "tail" of $f(E)$ extending into the conduction band of Fig. 3–15a is symmetrical with the hole probability tail $[1 - f(E)]$ in the valence band. The distribution function has values within the band gap between E_v and E_c, but there are no energy states available, and no electron occupancy results from $f(E)$ in this range.



(a) Intrinsic (b) n-type



(c) p-type

Figure 3–15 The Fermi distribution function applied to semiconductors: (a) intrinsic material; (b) n-type material; (c) p-type material.

---

[4] Actually the intrinsic E_F is displaced slightly from the middle of the gap, since the densities of available states in the valence and conduction bands are not equal (Section 3.3.2).

The tails in $f(E)$ are exaggerated in Fig. 3–15 for illustrative purposes. Actually, the probability values at E_v and E_c are quite small for intrinsic material at reasonable temperatures. For example, in Si at 300 K, $n_i = p_i \approx 10^{10}$ cm^{-3}, whereas the densities of available states at E_v and E_c are on the order of 10^{19} cm^{-3}. Thus the probability of occupancy $f(E)$ for an individual state in the conduction band and the hole probability $[1 - f(E)]$ for a state in the valence band are quite small. Because of the relatively large density of states in each band, small changes in $f(E)$ can result in significant changes in carrier concentration.

In n-type material there is a high concentration of electrons in the conduction band compared with the hole concentration in the valence band (recall Fig. 3–12a). Thus in n-type material the distribution function $f(E)$ must lie above its intrinsic position on the energy scale (Fig. 3–15b). Since $f(E)$ retains its shape for a particular temperature, the larger concentration of electrons at E_c in n-type material implies a correspondingly smaller hole concentration at E_v. We notice that the value of $f(E)$ for each energy level in the conduction band (and therefore the total electron concentration n_0) increases as E_F moves closer to E_c. Thus the energy difference $(E_c - E_F)$ gives a measure of n; we shall express this relation mathematically in the following section.

For p-type material the Fermi level lies near the valence band (Fig. 3–15c) such that the $[1 - f(E)]$ tail below E_v is larger than the $f(E)$ tail above E_c. The value of $(E_F - E_v)$ indicates how strongly p-type the material is.

It is usually inconvenient to draw $f(E)$ vs. E on every energy band diagram to indicate the electron and hole distributions. Therefore, it is common practice merely to indicate the position of E_F in band diagrams. This is sufficient information, since for a particular temperature the position of E_F implies the distributions in Fig. 3–15.

3.3.2 Electron and Hole Concentrations at Equilibrium

The Fermi distribution function can be used to calculate the concentrations of electrons and holes in a semiconductor, if the densities of available states in the valence and conduction bands are known. For example, the concentration of electrons in the conduction band is

$$n_0 = \int_{E_c}^{\infty} f(E)N(E)dE \qquad (3\text{–}12)$$

where $N(E)dE$ is the density of states (cm^{-3}) in the energy range dE. The subscript 0 used with the electron and hole concentration symbols (n_0, p_0) indicates equilibrium conditions. The number of electrons per unit volume in the energy range dE is the product of the density of states and the probability of occupancy $f(E)$. Thus the total electron concentration is the integral