1

Semiconductor Physics

2   Principles of Solar Cells, LEDs and Diodes

Objectives

1. Understand semiconductor band theory and its relevance to semiconductor devices.
2. Obtain a qualitative understanding of how bands depend on semiconductor materials.
3. Introduce the concept of the Fermi energy.
4. Introduce the concept of the mobile hole in semiconductors.
5. Derive the number of mobile electrons and holes in semiconductor bands.
6. Obtain expressions for the conductivity of semiconductor material based on the electron and hole concentrations and mobilities.
7. Introduce the concepts of doped semiconductors and the resulting electrical characteristics.
8. Understand the concept of excess, non-equilibrium carriers generated by either illumination or by current flow due to an external power supply.
9. Introduce the physics of traps and carrier recombination and generation.
10. Introduce alloy semiconductors and the distinction between direct gap and indirect gap semiconductors.

1.1   Introduction

A fundamental understanding of electron behaviour in crystalline solids is available using the *band theory of solids.* This theory explains a number of fundamental attributes of electrons in solids including:

 (i) concentrations of charge carriers in semiconductors;
 (ii) electrical conductivity in metals and semiconductors;
(iii) optical properties such as absorption and photoluminescence;
(iv) properties associated with junctions and surfaces of semiconductors and metals.

   The aim of this chapter is to present the theory of the band model, and then to exploit it to describe the important electronic properties of semiconductors. This is essential for a proper understanding of p-n junction devices, which constitute both the photovoltaic (PV) solar cell and the light-emitting diode (LED).

1.2   The Band Theory of Solids

There are several ways of explaining the existence of energy bands in crystalline solids. The simplest picture is to consider a single atom with its set of discrete energy levels for its electrons. The electrons occupy quantum states with quantum numbers n, l, m and s denoting the energy level, orbital and spin state of the electrons. Now if a number N of identical atoms are brought together in very close proximity as in a crystal, there is some degree of spatial overlap of the outer electron orbitals. This means that there is a chance that

any pair of these outer electrons from adjacent atoms could trade places. The Pauli exclusion principle, however, requires that each electron occupy a unique energy state. Satisfying the Pauli exclusion principle becomes an issue because electrons that trade places effectively occupy new, *spatially extended* energy states. The two electrons apparently occupy the same spatially extended energy state.

In fact, since outer electrons from all adjacent atoms may trade places, outer electrons from *all* the atoms may effectively trade places with each other and therefore a set of outermost electrons from the *N* atoms all appear to share a spatially extended energy state that extends through the entire crystal. The Pauli exclusion principle can only be satisfied if these electrons occupy a set of *distinct*, spatially extended energy states. This leads to a set of slightly different energy levels for the electrons that all originated from the same atomic orbital. We say that the atomic orbital splits into an *energy band* containing a set of electron states having a set of closely spaced energy levels. Additional energy bands will exist if there is some degree of spatial overlap of the atomic electrons in lower-lying atomic orbitals. This results in a set of energy bands in the crystal. Electrons in the lowest-lying atomic orbitals will remain virtually unaltered since there is virtually no spatial overlap of these electrons in the crystal.

The picture we have presented is conceptually a very useful one and it suggests that electrical conductivity may arise in a crystal due to the formation of spatially extended electron states. It does not directly allow us to quantify and understand important details of the behaviour of these electrons, however.

We need to understand the behaviour in a solid of the electrons that move about in the material. These mobile charge carriers are crucially important in terms of the electrical properties of devices. An electron inside an infinitely large vacuum chamber is a free electron, but a mobile electron in a solid behaves very differently.

We can obtain a more detailed model as follows. The mobile electrons in a crystalline semiconductor are influenced by the electric potential in the material. This potential has a spatial periodicity on an atomic scale due to the crystal structure. For example, positively charged atomic sites provide potential valleys to a mobile electron and negatively charged atomic sites provide potential peaks or barriers. In addition, the semiconductor is finite in its spatial dimensions and there will be additional potential barriers or potential changes at the boundaries of the semiconductor material.

The quantitative description of these spatially extended electrons requires the use of wavefunctions that include their spatial distribution as well as their energy and momentum. These wavefunctions may be obtained by solving Schrödinger's equation. The following section presents a very useful band theory of crystalline solids and the results.

1.3   The Kronig–Penney Model

The *Kronig–Penney model* is able to explain the essential features of band theory.

First, consider an electron that can travel within a one-dimensional periodic potential $V(x)$. The periodic potential can be considered as a series of regions having zero potential energy separated by potential energy barriers of height V_0, as shown in Figure 1.1, forming a simple periodic potential with period $a + b$. We associate $a + b$ also with the lattice constant of the crystal. Note that the electric potential in a real crystal does not exhibit the
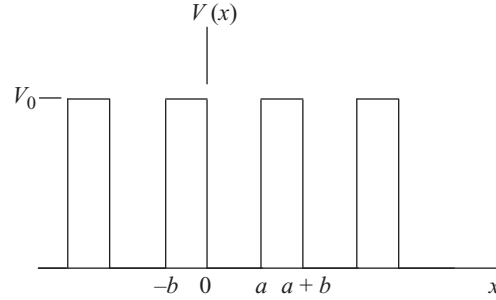
Figure 1.1 *Simple one-dimensional potential* V*(x) used in the Kronig–Penney model*

idealized shape of this periodic potential; however, the result turns out to be relevant in any case, and Schrödinger's equation is much easier to solve starting from the potential of Figure 1.1.

In order to obtain the electron wavefunctions relevant to an electron in the crystalline solid, $V(x)$ is substituted into the time-independent form of Schrödinger's equation:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + V(x) = E\psi(x) \tag{1.1}$$

where $V(x)$ is the potential energy and E is total energy.

For $0 \leq x \leq a$ we have $V = 0$ and the general solution to Equation 1.1 yields:

$$\psi(x) = Ae^{iKx} + Be^{-iKx} \tag{1.2a}$$

where

$$E = \frac{\hbar^2 K^2}{2m} \tag{1.2b}$$

For $-b \leq x \leq 0$ we have

$$\psi(x) = Ce^{Qx} + De^{-Qx} \tag{1.3a}$$

where

$$V_0 - E = \frac{\hbar^2 Q^2}{2m} \tag{1.3b}$$

Boundary conditions must be satisfied such that $\psi(x)$ and $\frac{d\psi(x)}{dx}$ are continuous functions. At $x = 0$, equating (1.2a) and (1.3a), we have

$$A + B = C + D \tag{1.4a}$$

and equating derivatives of (1.2a) and (1.3a),

$$iK(A - B) = Q(C - D) \tag{1.4b}$$

An important additional constraint on the required wavefunctions results from the periodicity of the lattice. The solution to Equation 1.1 for any periodic potential must also have the form of a *Bloch function*:

$$\psi(x) = u_k(x)e^{ikx} \tag{1.5}$$

Here, k is the wavenumber of a plane wave. There are no restrictions on this wavenumber; however, $u_k(x)$ must be a periodic function with the same periodicity as the lattice.

Consider two x-values separated by one lattice constant, namely $x = -b$ and $x = a$. Now, Equation 1.5 states that $\psi(x + a + b) = \psi(x)e^{ik(a+b)}$. At $x = -b$ this may be written as:

$$\psi(a) = \psi(-b)e^{ik(a+b)} \tag{1.6}$$

The boundary conditions to satisfy $\psi(x)$ and $\frac{d\psi(x)}{dx}$ being continuous functions at $x = a$ may now be written by substituting ψ from Equations 1.2 and 1.3 into Equation 1.6:

$$Ae^{iKa} + Be^{-iKa} = (Ce^{-Qb} + De^{Qb})e^{ik(a+b)} \tag{1.7a}$$

and substituting the corresponding derivatives:

$$iK(Ae^{iKa} - Be^{-iKa}) = Q(Ce^{-Qb} + De^{Qb})e^{ik(a+b)} \tag{1.7b}$$

Equations 1.4a, 1.4b, 1.7a and 1.7b constitute four equations with four unknowns A, B, C and D. A solution exists only if the determinant of the coefficients of A, B, C and D is zero (Cramer's rule). This requires that

$$\frac{Q^2 - K^2}{2QK} \sinh QB \sin Ka + \cosh Qb \cos Ka = \cos k(a + b) \tag{1.7c}$$

This may be simplified if the limit $b \rightarrow 0$ and $V_0 \rightarrow \infty$ is taken such that bV_0 is constant (see Problem 1.1). We now define

$$P = \frac{Q^2 ba}{2}$$

Since $Q \gg K$ and $Qb \ll 1$ we obtain

$$\cos ka = P \frac{\sin Ka}{Ka} + \cos Ka \tag{1.8}$$

Here k is the wavevector of the electron describing its momentum $p = \hbar k$ and

$$K = \frac{1}{\hbar}\sqrt{2mE} \tag{1.9}$$

which means that K is a term associated with the electron's energy.

Now, Equation 1.8 only has solutions if the righthand side of Equation 1.8 is between -1 and $+1$, which restricts the possible values of Ka. The righthand side is plotted as a function of Ka in Figure 1.2.
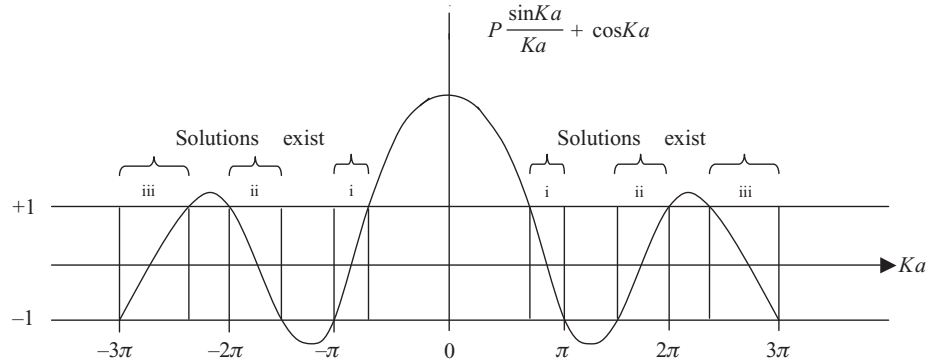
6   *Principles of Solar Cells, LEDs and Diodes*

Figure 1.2   *Graph of righthand side of Equation 1.8 as a function of* P *for* P = 2

Since *K* and *E* are related by Equation 1.9, these allowed ranges of *Ka* actually describe *energy bands* (allowed ranges of *E*) separated by *energy gaps* (forbidden ranges of *E*). *Ka* may be re-plotted on an energy axis, which is related to the *Ka* axis by the square root relationship of Equation 1.9. It is convenient to view *E* on a vertical axis as a variable dependent on *k*. Note that $k = \frac{n\pi}{a}$ for integer values of *n* at the edges of each energy band where the left side of Equation 1.8 is equal to ±1. These critical values of *k* occur at the boundaries of what are called *Brillouin zones*. A sketch of *E* versus *k* is shown in Figure 1.3, which clearly shows the energy bands and energy gaps.

Let us now plot the free electron graph for *E* versus *k*. Solving Equation 1.1 for a free electron with *V* = 0 yields the solution

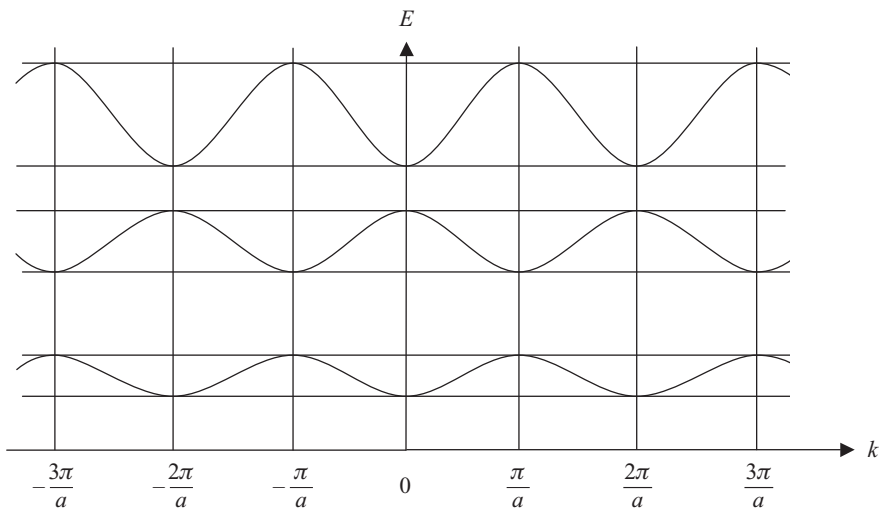$$\psi(x) = Ae^{ikx} + Be^{-ikx}$$

Figure 1.3   *Plot of* E *versus* k *showing how* k *varies within each energy band and the existence of energy bands and energy gaps. The vertical lines at* k = n$\frac{\pi}{a}$ *are Brillouin zone boundaries*
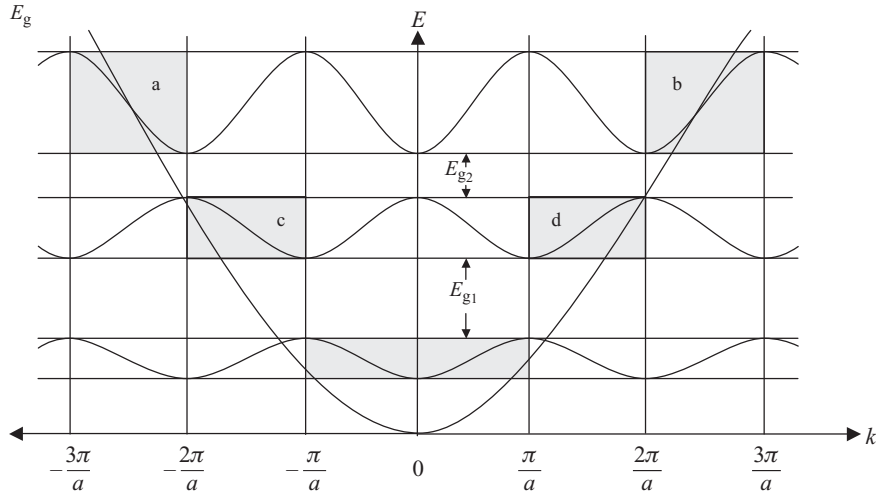
Figure 1.4 *Plot of* E *versus* k *comparing the result of the Kronig–Penney model to the free electron parabolic result*

where

$$E = \frac{\hbar^2 k^2}{2m} \tag{1.10}$$

This parabolic E versus k relationship is plotted superimposed on the curves from Figure 1.3. The result is shown in Figure 1.4.

Taking the limit $P \rightarrow 0$, and combining Equations 1.8 and 1.9, we obtain:

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

which is identical to Equation 1.10. This means that the dependence of E on k in Figure 1.4 approaches a parabola as expected if the amplitude of the periodic potential is reduced to zero. In fact, the relationship between the parabola and the Kronig–Penney model is evident if we look at the solutions to Equation 1.4 within the shaded regions in Figure 1.4 and regard them as portions of the parabola that have been broken up by energy gaps and distorted in shape. For a weak periodic potential (small P) the solutions to Equation 1.4 would more closely resemble the parabola. We refer to Equation 1.10 as a *dispersion relation* – it relates energy to the wavenumber of a particle.

At this point, we can draw some very useful conclusions based on the following result: *The size of the energy gaps increases as the periodic potential increases in amplitude in a crystalline solid.* Periodic potentials are larger in amplitude for crystalline semiconductors that have small atoms since there are then fewer atomically bound electrons to screen the point charges of the nuclei of the atoms. In addition, periodic potentials increase in amplitude for compound semiconductors as the ionic character of the crystal bonding increases. This will be illustrated in Section 1.10 for some real semiconductors.

To extend our understanding of energy bands we now need to turn to another picture of electron behaviour in a crystal.

1.4   The Bragg Model

Since electrons behave like waves, they will exhibit the behaviour of waves that undergo reflections. Notice that in a crystal with lattice constant a, the Brillouin zone boundaries occur at

$$k = \frac{n\pi}{a} = \frac{2\pi}{\lambda}$$

which may be rearranged to obtain

$$2a = n\lambda$$

The well-known Bragg condition relevant to waves that undergo strong reflections when incident on a crystal with lattice constant a is

$$2a \sin\theta = n\lambda$$

Now, if the electron is treated as a wave incident at $\theta = 90°$ then we have

$$2a = n\lambda$$

which is precisely the case at Brillouin zone boundaries. We therefore make the following observation: Brillouin zone boundaries occur when the electron wavelength satisfies the requirement for strong reflections from crystal lattice planes according to the Bragg condition. The free electron parabola in Figure 1.4 is similar to the Kronig–Penney model in the shaded regions well away from Brillouin zone boundaries; however, as we approach Brillouin zone boundaries, strong deviations take place and energy gaps are observed.

There is therefore a fundamental connection between the Bragg condition and the formation of energy gaps. The electrons that satisfy the Bragg condition actually exist as *standing waves* since reflections will occur equally for electrons travelling in both directions of the x axis, and standing waves do not travel. Provided electrons have wavelengths not close to the Bragg condition, they interact relatively weakly with the crystal lattice and behave more like free electrons.

The E versus k dependence immediately above and below any particular energy gap is contained in four shaded regions in Figure 1.4. For example, the relevant shaded regions for E_{g2} in Figure 1.4 are labelled a, b, c and d. These four regions are redrawn in Figure 1.5. Energy gap E_{g2} occurs at $k = \pm\frac{2\pi}{a}$. Since this is a standing wave condition with both electron velocity and electron momentum $p = \hbar k$ equal to zero, E_{g2} is redrawn at $k = 0$ in Figure 1.5. Since we are only interested in relative energies, the origin of the energy axis is moved for convenience, and we can arbitrarily redefine the origin of the energy axis. Figure 1.5 is known as a *reduced zone scheme*.

1.5   Effective Mass

We now introduce the concept of *effective mass* m^* to allow us to quantify electron behaviour. Effective mass changes in a peculiar fashion near Brillouin zone boundaries, and generally is not the same as the free electron mass m. It is easy to understand that the effective acceleration of an electron in a crystal due to an applied electric field will depend
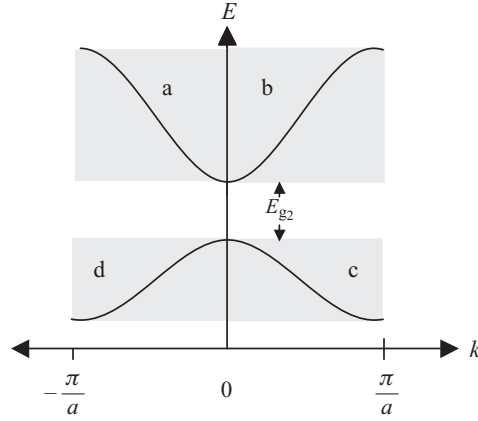
Figure 1.5 *Plot of* E *versus* k *in reduced zone scheme taken from regions a, b, c and d in Figure 1.4*

strongly on the nature of the reflections of electron waves off crystal planes. Rather than trying to calculate the specific reflections for each electron, we instead modify the mass of the electron to account for its observed willingness to accelerate in the presence of an applied force.

To calculate m^* we start with the free electron relationship

$$E = \frac{1}{2}mv_g^2$$

where v_g is the group velocity of the electron. Upon differentiation with respect to k,

$$\frac{dE}{dk} = mv_g\frac{dv_g}{dk} \tag{1.11}$$

Since $p = \hbar k = mv_g$ we can write

$$\frac{v_g}{k} = \frac{dv_g}{dk} = \frac{\hbar}{m} \tag{1.12}$$

Combining Equations 1.11 and 1.12 we obtain

$$\frac{dE}{dk} = v_g\hbar$$

or

$$v_g = \frac{1}{\hbar}\frac{dE}{dk} \tag{1.13}$$

Note that the group velocity falls to zero at the Brillouin zone boundaries where the slope of the E versus k graph is zero. This is consistent with the case of a standing wave.

Now, using Newton's law,

$$F = \frac{dp}{dt} = \hbar\frac{dk}{dt} \tag{1.14}$$

From Equations 1.13 and 1.14, we can write

$$\frac{\mathrm{d}v_g}{\mathrm{d}t} = \frac{1}{\hbar}\frac{\mathrm{d}^2 E}{\mathrm{d}k\mathrm{d}t} = \frac{1}{\hbar}\frac{\mathrm{d}^2 E}{\mathrm{d}k^2}\frac{\mathrm{d}k}{\mathrm{d}t} = \frac{F}{\hbar^2}\frac{\mathrm{d}^2 E}{\mathrm{d}k^2} \tag{1.15}$$

If we assign m^* to represent an effective electron mass, then Newton's law tells us that

$$\frac{\mathrm{d}v_g}{\mathrm{d}t} = \frac{F}{m^*}$$

Upon examination Equation 1.15 actually expresses Newton's law provided we define

$$m^* = \frac{\hbar^2}{\frac{d^2 E}{dk^2}} \tag{1.16}$$

Since $\frac{\mathrm{d}^2 E}{\mathrm{d}k^2}$ is the curvature of the plot in Figure 1.5, it is interesting to note that m^* will be negative for certain values of k. This may be understood physically: if an electron that is close to the Bragg condition is accelerated slightly by an applied force it may then move even closer to the Bragg condition, reflect more strongly off the lattice planes, and effectively accelerate in the direction opposite to the applied force.

We can apply Equation 1.16 to the free electron case where $E = \frac{\hbar^2 k^2}{2m}$ and we immediately see that $m^* = m$ as expected. In addition at the bottom or top of energy bands illustrated in Figure 1.5, the shape of the band may be approximated as parabolic for small values of k and hence a constant effective mass is often sufficient to describe electron behaviour for small values of k. This will be useful when we calculate the number of electrons in an energy band.

1.6   Number of States in a Band

The curves in Figure 1.5 are misleading in that electron states in real crystals are discrete and only a finite number of states exist within each energy band. This means that the curves should be regarded as closely spaced dots that represent quantum states. We can determine the number of states in a band by considering a semiconductor crystal of length L and modelling the crystal as an infinite-walled potential box of length L with a potential of zero inside the well. See Example 1.1.

Example 1.1

An electron is inside a potential box of length L with infinite walls and zero potential in the box. The box is shown below.
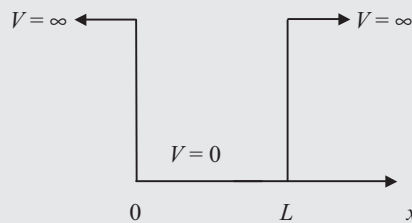
(a) Find the allowed energy levels in the box.
(b) Find the wavefunctions of these electrons.

Solution

(a) Inside the box, from Schrödinger's equation, we can substitute $V(x) = 0$ and we obtain

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} = E\psi(x)$$

Solutions are of the form

$$\psi(x) = A\exp\frac{i\sqrt{2mE}}{\hbar}x + B\exp\frac{-i\sqrt{2mE}}{\hbar}x$$

In regions where $V = \infty$ the wavefunction is zero. In order to avoid discontinuities in the wavefunction we satisfy boundary conditions at $x = 0$ and at $x = L$ and require that $\psi(0) = 0$ and $\psi(L) = 0$. These boundary conditions can be written

$$0 = A + B \text{ or } B = -A$$

and

$$0 = A\exp\frac{i\sqrt{2mE}}{\hbar}L + B\exp\frac{-i\sqrt{2mE}}{\hbar}L$$

$$= A\left(\exp\frac{i\sqrt{2mE}}{\hbar}L - \exp\frac{-i\sqrt{2mE}}{\hbar}L\right) = C\sin\frac{\sqrt{2mE}}{\hbar}L$$

where C is a constant. Now $\sin\theta$ is zero provided $\theta = n\pi$ where n is an integer and hence

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

A discrete set of allowed energy values is obtained by solving for E to obtain

$$E_n = \frac{n^2\pi^2\hbar^2}{2mL^2}$$

(b) The corresponding wavefunctions may be found by substituting the allowed energy values into Schrödinger's equation and solving:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} = \frac{n^2\pi^2\hbar^2}{2mL^2}\psi(x)$$

now

$$\frac{d^2\psi(x)}{dx^2} = -\frac{n^2\pi^2}{L^2}\psi(x)$$

and hence

$$\psi_n(x) = A \sin\left(\frac{n\pi}{L}x\right)$$

From Example 1.1 we obtain

$$\psi_n(x) = A \sin\frac{n\pi}{L}x \tag{1.17}$$

where n is a quantum number, and

$$k = \frac{n\pi}{L}, n = 1, 2, 3\ldots$$

As n increases we will inevitably reach the k value corresponding to the Brillouin zone boundary from the band model

$$k = \frac{\pi}{a}$$

This will occur when

$$\frac{n\pi}{L} = \frac{\pi}{a}$$

and therefore $n = \frac{L}{a}$. The maximum possible value of n now becomes the macroscopic length of the semiconductor crystal divided by the unit cell dimension, which is simply the number of unit cells in the crystal, which we shall call N. Since electrons have an additional quantum number s (spin quantum number) that may be either $\frac{1}{2}$ or $-\frac{1}{2}$, the maximum number of electrons that can occupy an energy band becomes

$$n = 2N$$

Although we have considered a one-dimensional model, the results can readily be extended into two or three dimensions and we still obtain the same result. See Problem 1.3.

We are now ready to determine the actual number of electrons in a band, which will allow us to understand electrical conductivity in semiconductor materials.

1.7   Band Filling

The existence of $2N$ electron states in a band does not determine the actual number of electrons in the band. At low temperatures, the electrons will occupy the lowest allowed energy levels, and in a semiconductor like silicon, which has 14 electrons per atom, several low-lying energy bands will be filled. In addition, the highest occupied energy band will be full, and then the next energy band will be empty. This occurs because silicon has an even number of valence electrons per unit cell, and when there are N unit cells, there will
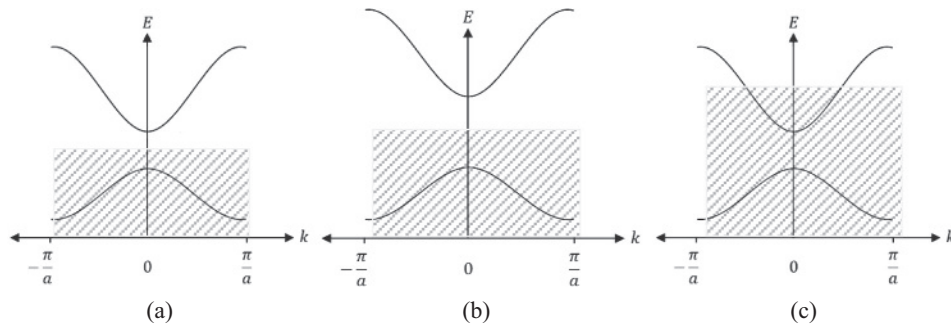
Figure 1.6 *The degree of filling of the energy bands in (a) semiconductors, (b) insulators and (c) metals at temperatures approaching 0 K. Available electron states in the hatched regions are filled with electrons and the energy states at higher energies are empty*

be the correct number of electrons to fill the $2N$ states in the highest occupied energy band. A similar argument occurs for germanium as well as carbon (diamond) although diamond is an insulator due to its large energy gap.

Compound semiconductors such as GaAs and other III-V semiconductors as well as CdS and other II-VI semiconductors exhibit the same result: The total number of electrons per unit cell is even, and at very low temperatures in a semiconductor the highest occupied band is filled and the next higher band is empty.

In many other crystalline solids this is not the case. For example group III elements Al, Ga and In have an odd number of electrons per unit cell, resulting in the highest occupied band being half filled since the $2N$ states in this band will only have N electrons to fill them. These are metals. Figure 1.6 illustrates the cases we have described, showing the electron filling picture in semiconductors, insulators and metals.

In Figure 1.6a the highest filled band is separated from the lowest empty band by an energy gap E_g that is typically in the range from less than 1 eV to between 3 and 4 eV in semiconductors. A completely filled energy band will not result in electrical conductivity because for each electron with positive momentum $p = \hbar k$ there will be one having negative momentum $p = -\hbar k$ resulting in no net electron momentum and hence no net electron flux even if an electric field is applied to the material.

Electrons may be promoted across the energy gap E_g by thermal energy or optical energy, in which case the filled band is no longer completely full and the empty band is no longer completely empty, and now electrical conduction occurs.

Above this range of E_g lie insulators (Figure 1.6b), which typically have an E_g in the range from about 4 eV to over 6 eV. In these materials it is difficult to promote electrons across the energy gap.

In metals, Figure 1.6c shows a partly filled energy band as the highest occupied band. The energy gap has almost no influence on electrical properties whereas occupied and vacant electron states within this partly filled band are significant: strong electron conduction takes place in metals because empty states exist in the highest occupied band, and electrons may be promoted very easily into higher energy states within this band. A very small applied electric field is enough to promote some electrons into higher energy states that impart a

14 *Principles of Solar Cells, LEDs and Diodes*

net momentum to the electrons within the band and an electron flow results, which results in the high electrical conductivity in metals.

1.8 Fermi Energy and Holes

Of particular interest is the existence in semiconductors, at moderate temperatures such as room temperature, of the two energy bands that are partly filled. The higher of these two bands is mostly empty but a number of electrons exist near the bottom of the band, and the band is named the *conduction band* because a net electron flux or flow may be obtained in this band. The lower band is almost full; however, because there are empty states near the top of this band, it also exhibits conduction and is named the *valence band*. The electrons that occupy it are valence electrons, which form covalent bonds in a semiconductor such as silicon. Figure 1.7 shows the room temperature picture of a semiconductor in thermal equilibrium. An imaginary horizontal line at energy E_f, called the *Fermi energy*, represents an energy above which the probability of electron states being filled is under 50%, and below which the probability of electron states being filled is over 50%. We call the empty states in the valence band *holes*. Both valence band holes and conduction band electrons contribute to conductivity.

In a semiconductor we can illustrate the valence band using Figure 1.8, which shows a simplified two-dimensional view of silicon atoms bonded covalently. Each covalent bond requires two electrons. The electrons in each bond are not unique to a given bond, and are shared between all the covalent bonds in the crystal, which means that the electron wavefunctions extend spatially throughout the crystal as described in the Kronig–Penney model. A valence electron can be thermally or optically excited and may leave a bond to form an *electron-hole pair* (EHP). The energy required for this is the bandgap energy of the
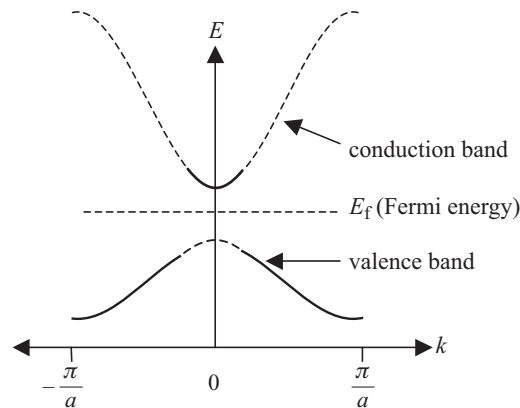

Figure 1.7 *Room temperature semiconductor showing the partial filling of the conduction band and partial emptying of the valence band. Valence band holes are formed due to electrons being promoted across the energy gap. The Fermi energy lies between the bands. Solid lines represent energy states that have a significant chance of being filled*
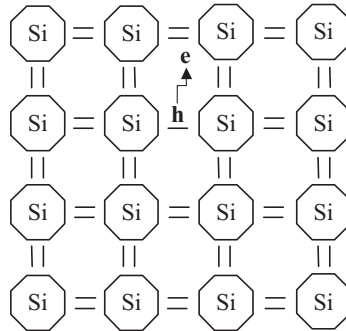
Figure 1.8 *Silicon atoms have four covalent bonds as shown. Although silicon bonds are tetrahedral, they are illustrated in two dimensions for simplicity. Each bond requires two electrons, and an electron may be excited across the energy gap to result in both a hole in the valence band and an electron in the conduction band that are free to move independently of each other*

semiconductor. Once the electron leaves a covalent bond a hole is created. Since valence electrons are shared, the hole is likewise shared among bonds and is able to move through the crystal. At the same time the electron that was excited enters the conduction band and is also able to move through the crystal resulting in two independent charge carriers.

In order to calculate the conductivity arising from a particular energy band, we need to know the number of electrons n per unit volume of semiconductor, and the number of holes p per unit volume of semiconductor resulting from the excitation of electrons across the energy gap E_g. In the special case of a pure or *intrinsic* semiconductor, we can write the carrier concentrations as n_i and p_i such that $n_i = p_i$

1.9   Carrier Concentration

The determination of n and p requires us to find the number of states in the band that have a significant probability of being occupied by an electron, and for each state we need to determine the probability of occupancy to give an appropriate weighting to the state.

We will assume a constant effective mass for the electrons or holes in a given energy band. In real semiconductor materials the relevant band states are either near the top of the valence band or near the bottom of the conduction band as illustrated in Figure 1.7. In both cases the band shape may be approximated by a parabola, which yields a constant curvature and hence a constant effective mass as expressed in Equation 1.16.

In contrast to effective mass, the probability of occupancy by an electron in each energy state depends strongly on energy, and we cannot assume a fixed value. We use the Fermi–Dirac distribution function, which may be derived from Boltzmann statistics as follows. Consider a crystal lattice having lattice vibrations, or *phonons*, that transfer energy to electrons in the crystal. These electrons occupy quantum states that can also transfer energy back to the lattice, and a thermal equilibrium will be established.

Consider an electron in a crystal that may occupy lower and higher energy states E_1^e and E_2^e respectively, and a lattice phonon that may occupy lower and higher energy states E_1^p and E_2^p respectively. Assume this electron makes a transition from energy E_1^e to E_2^e by accepting energy from the lattice phonon while the phonon makes a transition from E_2^p to E_1^p. For conservation of energy,

$$E_2^e - E_1^e = E_2^p - E_1^p \tag{1.18}$$

The probability of these transitions occurring can now be analysed. Let $p(E^e)$ be the probability that the electron occupies a state having energy E^e. Let $p(E^p)$ be the probability that the phonon occupies an energy state having energy E^p. For a system in thermal equilibrium the probability of an electron transition from E_1^e to E_2^e is the same as the probability of a transition from E_2^e to E_1^e, and we can write

$$p(E_2^p)p(E_1^e)(1 - p(E_2^e)) = p(E_1^p)p(E_2^e)(1 - p(E_1^e)) \tag{1.19}$$

because the probability that an electron makes a transition from E_1^e to E_2^e is proportional to the terms on the lefthand side in which the phonon at E_2^p must be available and the electron at E_1^e must be available. In addition, the electron state at E_2^e must be vacant because electrons, unlike phonons, must obey the Pauli exclusion principle, which allows only one electron per quantum state. Similarly the probability that the electron makes a transition from E_2^e to E_1^e is proportional to the terms on the righthand side.

From Boltzmann statistics (see Appendix 3) for phonons or lattice vibrations we use the Boltzmann distribution function:

$$p(E) \propto \exp\left(-\frac{E}{kT}\right) \tag{1.20}$$

Combining Equations 1.19 and 1.20 we obtain

$$\exp\left(-\frac{E_2^p}{kT}\right) p(E_1^e)(1 - p(E_2^e)) = \exp\left(-\frac{E_1^p}{kT}\right) p(E_2^e)(1 - p(E_1^e))$$

which may be written

$$p(E_1^e)(1 - p(E_2^e)) = \exp\left(\frac{E_2^p - E_1^p}{kT}\right) p(E_2^e)(1 - p(E_1^e))$$

Using Equation 1.18 this can be expressed entirely in terms of electron energy levels as

$$p(E_1^e)(1 - p(E_2^e)) = \exp\left(\frac{E_2^e - E_1^e}{kT}\right) p(E_2^e)(1 - p(E_1^e))$$

Rearranging this we obtain

$$\frac{p(E_1^e)}{1 - p(E_1^e)} \exp\left(\frac{E_1^e}{kT}\right) = \frac{p(E_2^e)}{1 - p(E_2^e)} \exp\left(\frac{E_2^e}{kT}\right) \tag{1.21}$$

The left side of this equation is a function only of the initial electron energy level and the right side is only a function of the final electron energy level. Since the equation must always hold and the initial and final energies may be chosen arbitrarily we must conclude that both sides of the equation are equal to an energy-independent quantity, which can only

be a function of the remaining variable T. Let this quantity be $f(T)$. Hence using either the left side or the right side of the equation we can write

$$\frac{p(E)}{1 - p(E)} \exp\left(\frac{E}{kT}\right) = f(T)$$

where E represents the electron energy level.

Solving for $p(E)$ we obtain

$$p(E) = \frac{1}{1 + \frac{1}{f(T)} \exp\left(\frac{E}{kT}\right)} \tag{1.22}$$

We now formally define the Fermi energy E_f to be the energy level at which $p(E) = \frac{1}{2}$ and hence

$$\frac{1}{f(T)} \exp\left(\frac{E_f}{kT}\right) = 1$$

or

$$\frac{1}{f(T)} = \exp\left(\frac{-E_f}{kT}\right)$$

Under equilibrium conditions the final form of the probability of occupancy at temperature T for an electron state having energy E is now obtained by substituting this into Equation 1.22 to obtain

$$F(E) = \frac{1}{1 + \exp\left(\frac{E - E_f}{kT}\right)} \tag{1.23}$$

where $F(E)$ is used in place of $p(E)$ to indicate that this is the Fermi–Dirac distribution function. This function is graphed in Figure 1.9.
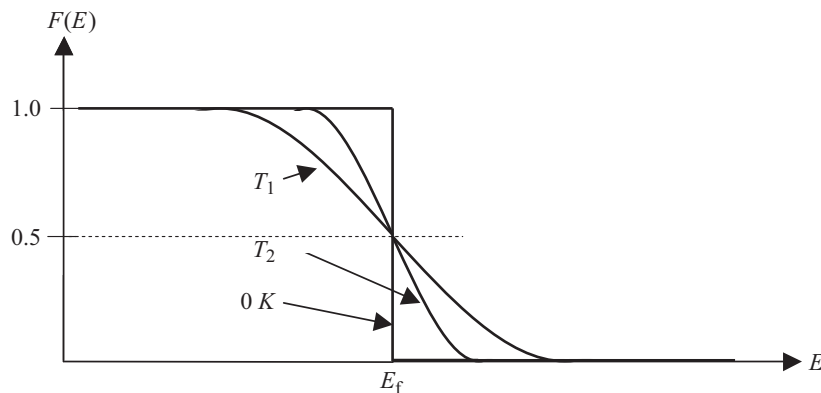

Figure 1.9 *Plot of the Fermi–Dirac distribution function F(E), which gives the probability of occupancy by an electron of an energy state having energy E. The plot is shown for two temperatures $T_1 > T_2$ as well as for 0 K. At absolute zero, the function becomes a step function*
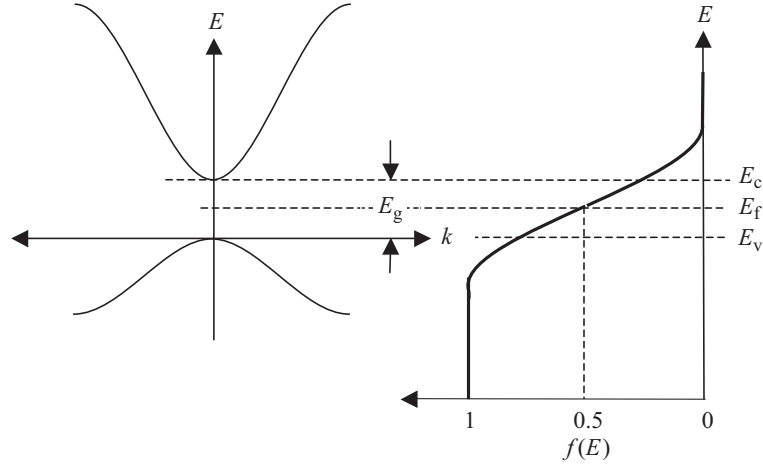
Figure 1.10   *A semiconductor band diagram is plotted along with the Fermi–Dirac distribution function. This shows the probability of occupancy of electron states in the conduction band as well as the valence band. Hole energies increase in the negative direction along the energy axis. The hole having the lowest possible energy occurs at the top of the valence band. This occurs because by convention the energy axis represents electron energies and not hole energies. The origin of the energy axis is located at E_v for convenience*

$F(E)$ is 0.5 at $E = E_f$ provided $T > 0$ K, and at high temperatures the transition becomes more gradual due to increased thermal activation of electrons from lower energy levels to higher energy levels. Figure 1.10 shows $F(E)$ plotted beside a semiconductor band diagram with the energy axis in the vertical direction. The bottom of the conduction band is at E_c and the top of the valence band is at E_v. At E_f there are no electron states since it is in the energy gap; however, above E_c and below E_v the values of $F(E)$ indicate the probability of electron occupancy in the bands. In the valence band the probability for a hole to exist at any energy level is $1 - F(E)$.

The distribution of available energy levels in an energy band is found by knowing the *density of states* function $D(E)$, which gives the number of available energy states per unit volume over a differential energy range. It is needed in order to calculate the number of carriers in an energy band. Knowing the probability of occupancy of the states in a band is not sufficient; the density of available energy levels is also required. Once we have all this information we can obtain the total number of electrons in a band by summing up all the electrons in each energy level within the energy band. The probability of each energy level being occupied by an electron is taken into account in the summation.

The density of states function may be derived by solving Schrödinger's equation for an infinite-walled potential box in which the wavefunctions (Equation 1.17) must be expressed in three dimensions. In three dimensions, Schrödinger's equation is:

$$-\frac{\hbar^2}{2m}\left(\frac{d^2}{dx^2} + \frac{d^2}{dy^2} + \frac{d^2}{dz^2}\right)\psi(x, y, z) + V(x, y, z)\psi(x, y, z) = E\psi(x, y, z)$$

Consider a box of dimensions a, b and c in three-dimensional space in which $V = 0$ inside the box when $0 < x < a, 0 < y < b, 0 < z < c$. Outside the box, assume $V = \infty$.

Inside the box using Schrödinger's equation:

$$-\frac{\hbar^2}{2m}\left(\frac{\mathrm{d}^2}{\mathrm{d}x^2}+\frac{\mathrm{d}^2}{\mathrm{d}y^2}+\frac{\mathrm{d}^2}{\mathrm{d}z^2}\right)\psi(x,y,z)=E\psi(x,y,z) \tag{1.24}$$

If we let $\psi(x,y,z)=X(x)Y(y)Z(z)$ then upon substitution into Equation 1.24 and after dividing by $\psi(x,y,z)$ we obtain:

$$-\frac{\hbar^2}{2m}\left(\frac{1}{X(x)}\frac{\mathrm{d}^2X(x)}{\mathrm{d}x^2}+\frac{1}{Y(y)}\frac{\mathrm{d}^2Y(y)}{\mathrm{d}y^2}+\frac{1}{Z(z)}\frac{\mathrm{d}^2Z(z)}{\mathrm{d}z^2}\right)=E$$

Since each term contains an independent variable, we can apply *separation of variables* and conclude that each term is equal to a constant that is independent of x, y and z.

Now, we have three equations

$$\frac{1}{X(x)}\frac{\mathrm{d}^2X(x)}{\mathrm{d}x^2}=-C_1 \tag{1.25a}$$

$$\frac{1}{Y(y)}\frac{\mathrm{d}^2Y(y)}{\mathrm{d}y^2}=-C_2 \tag{1.25b}$$

and

$$\frac{1}{Z(z)}\frac{\mathrm{d}^2Z(z)}{\mathrm{d}z^2}=-C_3 \tag{1.25c}$$

where

$$E=\frac{\hbar^2}{2m}(C_1+C_2+C_3) \tag{1.26}$$

The general solution to Equation 1.25a is

$$X(x)=A_1\exp(ikx)+A_2\exp(-ikx) \tag{1.27}$$

To satisfy boundary conditions such that $X(x)=0$ at $x=0$ and at $x=a$ we obtain

$$X(x)=A\sin k_x x$$

where

$$k_x=\frac{n_x\pi}{a}$$

with n_x a positive integer quantum number and

$$C_1=\left(\frac{n_x\pi}{a}\right)^2$$

Repeating a similar procedure for Equations 1.25b and 1.25c, and using Equation 1.26 we obtain:

$$\psi(x,y,z)=X(x)Y(y)Z(z)=ABC\sin(k_x x)\sin(k_y y)\sin(k_z z)$$

and

$$E=\frac{\hbar^2\pi^2}{2m}\left(\left(\frac{n_x}{a}\right)^2+\left(\frac{n_y}{b}\right)^2+\left(\frac{n_z}{c}\right)^2\right) \tag{1.28}$$
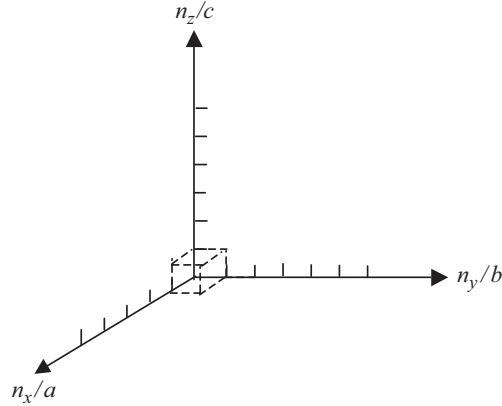
Figure 1.11 *Reciprocal space lattice. A cell in this space is shown, which is the volume associated with one lattice point. The cell has dimensions $\frac{1}{a}$, $\frac{1}{b}$, $\frac{1}{c}$ and volume $\frac{1}{abc}$. This space may readily be transformed into* k*-space by multiplying each axis by a factor of 2π. It may also readily be transformed into* p*-space (momentum space) by multiplying each axis by a factor of* h

If more than one electron is put into the box at zero kelvin the available energy states will be filled such that the lowest energy states are filled first.

We now need to determine how many electrons can occupy a specific energy range in the box. It is very helpful to define a three-dimensional space with coordinates $\frac{n_x}{a}$, $\frac{n_y}{b}$ and $\frac{n_z}{c}$. In this three-dimensional space there are discrete points that are defined by these coordinates with integer values of n_x, n_y and n_z in what is referred to as a *reciprocal space lattice*, which is shown in Figure 1.11. Note that reciprocal space is related to another important conceptual space known as *k-space* by multiplying each axis of reciprocal space by a factor of 2π.

From Equation 1.28 it is seen that an ellipsoidal shell in reciprocal space represents an equal energy surface because the general form of this equation is that of an ellipsoid in reciprocal space. The number of reciprocal lattice points that are contained inside the positive octant of an ellipsoid having a volume corresponding to a specific energy E will be the number of states smaller than E. The number of electrons is actually twice the number of these points because electrons have an additional quantum number s for spin and $s = \pm\frac{1}{2}$. The positive octant of the ellipsoid is illustrated in Figure 1.12.

Rearranging Equation 1.28,

$$\left(\frac{n_x}{a}\right)^2 + \left(\frac{n_y}{b}\right)^2 + \left(\frac{n_z}{c}\right)^2 = \frac{2mE}{\hbar^2\pi^2} \tag{1.29}$$

The number of reciprocal lattice points inside the ellipsoid is the volume of the ellipsoid divided by the volume associated with each lattice point shown in Figure 1.11. The volume of the ellipsoid is

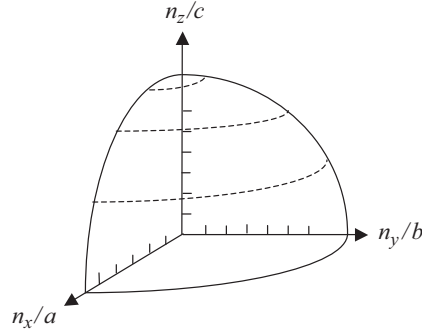$$V = \frac{4}{3}\pi \text{ (product of semi-axes of ellipsoid)}.$$

Figure 1.12 *The positive octant of an ellipsoid of revolution in reciprocal space corresponding to an equal energy surface. The number of electron states below this energy is twice the number of reciprocal lattice points inside the positive octant of the ellipsoid*

The semi-axis of length L_x along the $\frac{n_x}{a}$ direction is obtained by setting $n_y = n_z = 0$ and solving for $\frac{n_x}{a}$ in Equation 1.29, and we obtain

$$L_x = \sqrt{\frac{2mE}{\hbar^2 \pi^2}}$$

By repeating this for each semi-axis we obtain:

$$V = \frac{4}{3}\pi \left(\frac{2mE}{\hbar^2 \pi^2}\right)^{\frac{3}{2}}$$

Now if the volume of the ellipsoid is much larger than the volume associated with one lattice point then, including spin, the number of electrons having energy less than E approaches two times one-eighth of the volume of the ellipsoid (positive octant) divided by the volume associated with one lattice point, or:

$$number\ of\ electrons = \frac{2\left(\frac{1}{8}\right)\frac{4}{3}\pi \left(\frac{2mE}{\hbar^2 \pi^2}\right)^{\frac{3}{2}}}{\frac{1}{abc}}$$

We define $n(E)$ to be the number of electrons *per unit volume* of the box and therefore

$$n(E) = 2\left(\frac{1}{8}\right)\frac{4}{3}\pi \left(\frac{2mE}{\hbar^2 \pi^2}\right)^{\frac{3}{2}}$$

We also define $D(E)$ to be the *density of states* function where

$$D(E) = \frac{\mathrm{d}n(E)}{\mathrm{d}(E)}$$

and finally we obtain

$$D(E) = \frac{\pi}{2} \left(\frac{2m}{\hbar^2 \pi^2} \right)^{\frac{3}{2}} E^{\frac{1}{2}} \tag{1.30}$$

This form of the density of states function is valid for a box having $V = 0$ inside the box. In an energy band, however, V is a periodic function and the density of states function must be modified. This is easy to do because rather than the parabolic E versus k dispersion relation (Equation 1.10) for free electrons in which the electron mass is m, we simply use the E versus k dependence for an electron near the bottom or top of an energy band as illustrated in Figure 1.7, which may be approximated as parabolic for small values of k but using the appropriate effective mass. It is important to remember that the density of states function is based on a density of available states in k-space or reciprocal space, and that for a certain range of k-values in k-space the corresponding range of energies along the energy axis is determined by the slope of the E versus k graph. The slope of E versus k in a parabolic band depends on the effective mass (see Section 1.5). Using Equation 1.10 the relevant dispersion curve is now

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

As a result

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

and the second derivative or curvature of the E versus k curve is constant as required.

As a result the density of states function in a conduction band is given by Equation 1.30, provided the effective mass m^* is used in place of m. The point $E = 0$ should refer to the bottom of the band. We now have

$$D(E) = \frac{\pi}{2} \left(\frac{2m^*}{\hbar^2 \pi^2} \right)^{\frac{3}{2}} E^{\frac{1}{2}} \tag{1.31a}$$

Since E_v is defined as zero as in Figure 1.10 for convenience then the conduction band starts at $E_\mathrm{c} = E_\mathrm{g}$ and $D(E - E_\mathrm{g})$ tells us the number of energy states available per differential range of energy within the conduction band, and we obtain

$$D(E - E_\mathrm{g}) = \frac{\pi}{2} \left(\frac{2m^*_\mathrm{e}}{\hbar^2 \pi^2} \right)^{\frac{3}{2}} (E - E_\mathrm{g})^{\frac{1}{2}} \tag{1.31b}$$

The total number of electrons per unit volume in the band is now given by

$$n = \int_{E_\mathrm{g}}^{E_\mathrm{max}} D(E - E_\mathrm{g}) F(E) \mathrm{d}E \tag{1.32}$$

where E_max is the highest energy level in the energy band that needs to be considered as higher energy levels have a negligible chance of being occupied.

The integral may be solved analytically provided the upper limit of the integral is allowed to be infinity. This is justifiable because the argument of the integrand is virtually zero above E_max.

From Equation 1.23, since $E \geq E_g$ and $E_g - E_f \gg kT$, we can use the *Boltzmann approximation*:

$$F(E) \simeq \exp\left[\frac{-(E - E_f)}{kT}\right] \tag{1.33}$$

Hence from Equations 1.31, 1.32 and 1.33,

$$n = \frac{\pi}{2}\left(\frac{2m_e^*}{\hbar^2\pi^2}\right)^{\frac{3}{2}} \exp\left[-\frac{E_f}{kT}\right] \int_{E_g}^{E_{max}} (E - E_g)^{\frac{1}{2}} \exp\left[-\frac{E}{kT}\right] dE$$

$$= \frac{\pi}{2}\left(\frac{2m_e^*}{\hbar^2\pi^2}\right)^{\frac{3}{2}} \exp\left[-\frac{E_g - E_f}{kT}\right] \int_0^{\infty} (E)^{\frac{1}{2}} \exp\left[-\frac{E}{kT}\right] dE$$

From standard integral tables and because $E_c = E_g$ we obtain

$$n_0 = N_c \exp\left(\frac{-(E_c - E_f)}{kT}\right) \tag{1.34a}$$

where

$$N_c = 2\left(\frac{2\pi m_e^* kT}{h^2}\right)^{\frac{3}{2}} \tag{1.34b}$$

Here m_e^* is the effective mass of electrons near the bottom of the conduction band. The subscript on n indicates that equilibrium conditions apply. The validity of Equation 1.34 is maintained regardless of the choice of the origin on the energy axis since from Equation 1.34 the important quantity for determining the electron concentration is the *energy difference* between the conduction band edge and the Fermi energy.

The same procedure may be applied to the valence band. In this case we calculate the number of holes p in the valence band. The density of states function must be written as $D(-E)$ since from Figure 1.10 energy E is negative in the valence band and hole energy increases as we move in the negative direction along the energy axis. We can use Equation 1.31a to obtain

$$D(-E) = \frac{\pi}{2}\left(\frac{2m^*}{\hbar^2\pi^2}\right)^{\frac{3}{2}} (-E)^{\frac{1}{2}}$$

The probability of the existence of a hole is $1 - F(E)$, and from Equation 1.23 if $E_f - E \gg kT$ we obtain

$$1 - F(E) \cong \exp\left(\frac{E - E_f}{kT}\right)$$

and now

$$p = \int_0^{-E_{max}} D(-E)(1 - F(E)) dE$$

In an analogous manner to that described for the conduction band, we therefore obtain

$$p_0 = N_v \exp\left(\frac{-(E_f - E_v)}{kT}\right) \tag{1.35a}$$

where

$$N_v = 2 \left(\frac{2\pi m_h^* kT}{h^2} \right)^{\frac{3}{2}} \tag{1.35b}$$

and m_h^*, the hole effective mass, is a positive quantity.

Equation 1.35 shows that the important quantity for the calculation of hole concentration in Equation 1.16 is the *energy difference* between the Fermi energy and the valence band edge. Again the subscript on p indicates that equilibrium conditions apply.

We can now determine the position of the Fermi level and will again set $E_v = 0$ for convenience as illustrated in Figure 1.10. Since $n_i = p_i$ for an intrinsic semiconductor we equate Equations 1.34 and 1.35 and obtain

$$N_c \exp \left(\frac{-(E_g - E_f)}{kT} \right) = N_v \exp \left(\frac{-E_f}{kT} \right)$$

or

$$E_f = \frac{E_g}{2} + \frac{kT}{2} \ln \frac{N_v}{N_c} \tag{1.36}$$

The second term on the right side of Equation 1.36 is generally much smaller than $\frac{E_g}{2}$ and we conclude that the Fermi energy lies approximately in the middle of the energy gap.

From Equations 1.34 and 1.35 we can also write the product

$$np = N_c N_v \exp \left(\frac{-E_g}{kT} \right) \tag{1.37a}$$

and for an intrinsic semiconductor with $n_i = p_i$

$$n_i = p_i = \sqrt{N_c N_v} \exp \left(\frac{-E_g}{2kT} \right) \tag{1.37b}$$

which is a useful expression for carrier concentration as it is independent of E_f.

Example 1.2

(a) Calculate $n_i = p_i$ for silicon at room temperature and compare with the commonly accepted value.
(b) Calculate $n_i = p_i$ for gallium arsenide at room temperature.

Solution

(a) Using Appendix 2 to obtain silicon values $m_e^* = 1.08m$ and $E_g = 1.11$ eV,

$$N_c = 2 \left(\frac{2\pi m_e^* kT}{h^2} \right)^{\frac{3}{2}} = 2 \left(\frac{2\pi \times (1.08 \times 9.11 \times 10^{-31} \text{ kg}) \times (0.026 \times 1.6 \times 10^{-19} \text{ J})}{(6.625 \times 10^{-34} \text{ J s})^2} \right)^{\frac{3}{2}}$$
$$= 2.84 \times 10^{25} \text{ m}^{-3} = 2.84 \times 10^{19} \text{ cm}^{-3}$$

and

$$N_v = 2 \left(\frac{2\pi m_h^* kT}{h^2} \right)^{\frac{3}{2}} = 2 \left(\frac{2\pi \times (0.56 \times 9.11 \times 10^{-31} \text{ kg}) \times (0.026 \times 1.6 \times 10^{-19}\text{J})}{(6.625 \times 10^{-34}\text{J s})^2} \right)^{\frac{3}{2}}$$

$$= 1.06 \times 10^{25} \text{ m}^{-3} = 1.06 \times 10^{19} \text{ cm}^{-3}$$

Now,

$$n_i = p_i = \sqrt{N_c N_v} \exp \left(\frac{-E_g}{2kT} \right) = \sqrt{2.84 \times 10^{19} \times 1.06 \times 10^{19}}$$

$$\times \exp \left(\frac{-1.11 \text{ eV}}{2 \times 0.026 \text{ eV}} \right) = 9.31 \times 10^9 \text{cm}^{-3}$$

The commonly accepted value is $n_i = p_i = 1.5 \times 10^{10} \text{ cm}^{-3}$. The discrepancy relates mainly to three-dimensional aspects of the effective mass value, and the method and temperature at which effective mass is measured. We will continue to use the commonly accepted effective mass unless otherwise noted.

(b) For GaAs from Appendix 2, $m_e^* = 0.067$ and $E_g = 1.42$ eV. Hence

$$N_c = 2 \left(\frac{2\pi m_e^* kT}{h^2} \right)^{\frac{3}{2}} = 2 \left(\frac{2\pi \times (0.067 \times 9.11 \times 10^{-31} \text{ kg}) \times (0.026 \times 1.6 \times 10^{-19} \text{ J})}{(6.625 \times 10^{-34} \text{ J s})^2} \right)^{\frac{3}{2}}$$

$$= 4.38 \times 10^{23} \text{ m}^{-3} = 4.38 \times 10^{17} \text{ cm}^{-3}$$

and

$$N_v = 2 \left(\frac{2\pi m_h^* kT}{h^2} \right)^{\frac{3}{2}} = 2 \left(\frac{2\pi \times (0.48 \times 9.11 \times 10^{-31} \text{ kg}) \times (0.026 \times 1.6 \times 10^{-19} \text{ J})}{(6.625 \times 10^{-34} \text{ J s})^2} \right)^{\frac{3}{2}}$$

$$= 8.4 \times 10^{24} \text{ m}^{-3} = 8.4 \times 10^{18} \text{ cm}^{-3}$$

Now,

$$n_i = p_i = \sqrt{N_c N_v} \exp \left(\frac{-E_g}{2kT} \right) = \sqrt{4.38 \times 10^{17} \times 8.4 \times 10^{18}} \exp \left(\frac{-1.42 \text{ eV}}{2 \times 0.026 \text{ eV}} \right)$$

$$= 2.65 \times 10^6 \text{ cm}^{-3}$$

1.10 Semiconductor Materials

The relationship between carrier concentration and E_g has now been established and we can look at examples of real semiconductors. A portion of the periodic table showing elements from which many important semiconductors are made is shown in Figure 1.13, together with a list of selected semiconductors and their energy gaps. Note that there are the group IV semiconductors silicon and germanium, a number of III-V compound semiconductors

Group	II	III	IV	V	VI
Element		B	C	N	O
		Al	Si	P	S
	Zn	Ga	Ge	As	Se
	Cd	In	Sn	Sb	Te

Group	IV	IV	IV	IV	III-V	III-V	III-V	III-V	III-V	III-V	II-VI	II-VI
Element(s)	C	Si	Ge	Sn	GaN	AlP	GaP	AlAs	GaAs	InSb	ZnSe	CdTe
Energy gap (eV)	6	1.11	0.67	0	3.4	2.45	2.26	2.16	1.43	0.18	2.7	1.58

Figure 1.13   *A portion of the periodic table containing some selected semiconductors composed of elements in groups II to VI*

having two elements, one from group III and one from group V respectively, and a number of II-VI compound semiconductors having elements from group II and group VI respectively.

A number of interesting observations may now be made. In group IV crystals, the bonding is purely covalent. Carbon (diamond) is an insulator because it has an energy gap of 6 eV. The *energy gap decreases with atomic size* as we look down the group IV column from C to Si to Ge and to Sn. Actually Sn behaves like a metal. Since its energy gap is very small, it turns out that the valence band and conduction band effectively overlap when a three-dimensional model of the crystal is considered rather than the one-dimensional model we have discussed. This guarantees some filled states in the conduction band and empty states in the valence band regardless of temperature. Sn is properly referred to as a *semi-metal* (its conductivity is considerably lower than metals like copper or silver). We can understand this group IV trend of decreasing energy gaps since the periodic potential of heavy elements will be weaker than that of lighter elements due to electron screening as described in Section 1.3.

As with group IV materials, the energy gaps of III-V semiconductors decrease as we go down the periodic table from AlP to GaP to AlAs to GaAs and to InSb. The energy gaps of II-VI semiconductors behave in the same manner as illustrated by ZnSe and CdTe. Again, electron screening increases for heavier elements.

If we compare the energy gaps of a set of semiconductors composed of elements from the same row of the periodic table but with increasingly ionic bonding such as Ge, GaAs and ZnSe, another trend becomes clear: *Energy gaps increase as the degree of ionic character becomes stronger.* The degree of ionic bond character increases the magnitude of the periodic potential and hence the energy gap.

The carrier concentration as a function of temperature according to Equation 1.18 is plotted for three semiconductors in Figure 1.14. Increasing energy gaps result in lower carrier concentrations at a given temperature.

1.11   Semiconductor Band Diagrams

The semiconductors in Figure 1.10 crystallize in either cubic or hexagonal structures. Figure 1.15a shows the *diamond* structure of silicon, germanium (and carbon), which is cubic.
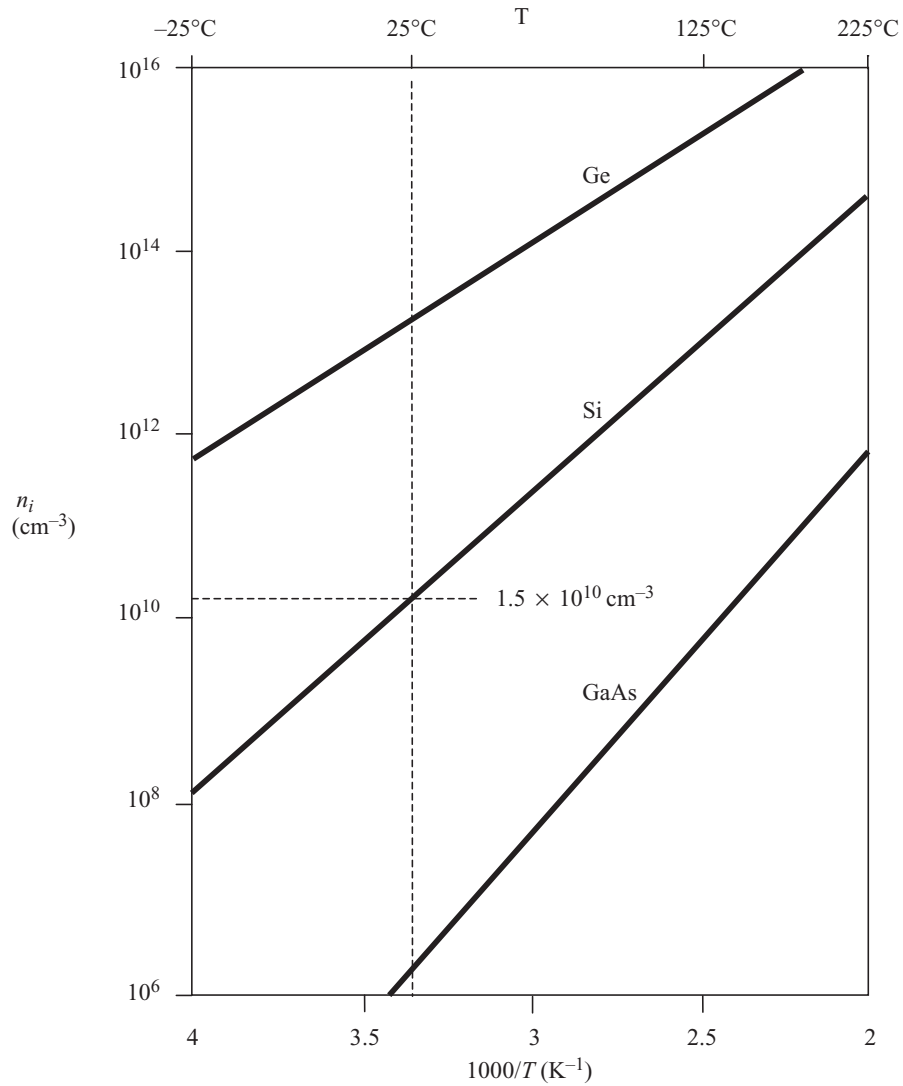
Figure 1.14 *Plot of commonly accepted values of* n *as a function of* $\frac{1}{T}$ *for intrinsic germanium* $(E_g = 0.7\ eV)$*, silicon* $(E_g = 1.1\ eV)$*, and gallium arsenide* $(E_g = 1.43\ eV)$

Figure 1.15b shows the *zincblende* structure of a set of III-V and II-VI semiconductors, which is also cubic. Figure 1.15c shows the *hexagonal* structure of some additional compound semiconductors.

These three structures have features in common. Each atom has four nearest neighbours in a tetrahedral arrangement. Some crystals exhibit distortions from the ideal 109.47-degree tetrahedral bond angle; however, since all the compounds have directional covalent bonding
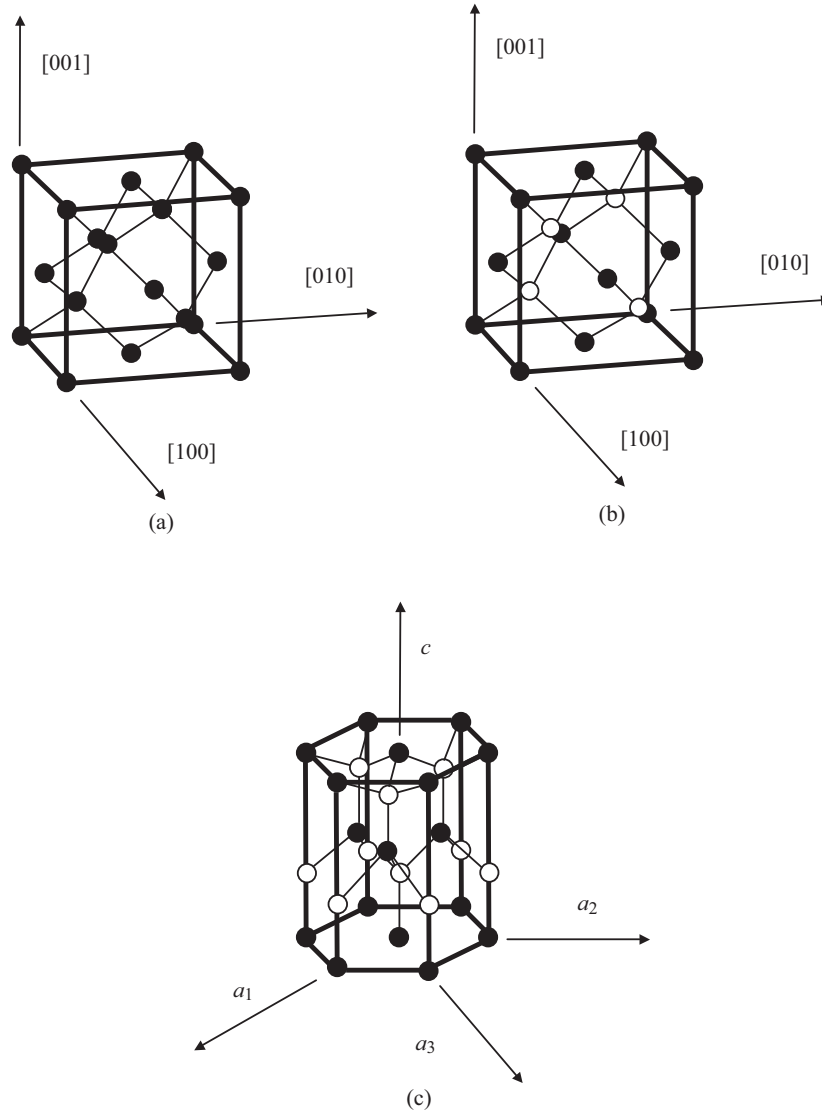
Figure 1.15 *(a) The diamond unit cell of crystal structures of C, Si and Ge. The cubic unit cell contains eight atoms. Each atom has four nearest neighbours in a tetrahedral arrangement. Within each unit cell, four atoms are arranged at the cube corners and at the face centres in a face-centred cubic (FCC) sublattice, and the other four atoms are arranged in another FCC sublattice that is offset by a translation along one quarter of the body diagonal of the unit cell. (b) The zincblende unit cell contains four 'A' atoms (black) and four 'B' atoms (white). The 'A' atoms form an FCC sublattice and the 'B' atoms form another FCC sublattice that is offset by a translation along one quarter of the body diagonal of the unit cell. (c) The hexagonal wurtzite unit cell contains six 'A' atoms and six 'B' atoms. The 'A' atoms form a hexagonal close-packed (HCP) sublattice and the 'B' atoms form another HCP sublattice that is offset by a translation along the vertical axis of the hexagonal unit cell. Each atom is tetrahedrally bonded to four nearest neighbours. A vertical axis in the unit cell is called the* c-axis

to some degree, bond angles do not vary widely. Both the cubic (111) planes and the wurtzite (1000) planes normal to the *c*-axis have close-packed hexagonal atomic arrangements.

The energy gap and effective mass values for a given semiconductor are not sufficient information for optoelectronic applications. We need to re-examine the energy band diagrams for real materials in more detail.

The Kronig–Penney model involves several approximations. A one-dimensional periodic potential instead of a three-dimensional periodic potential is used. The periodic potential is simplified, and does not actually replicate the atomic potentials in real semiconductor crystals. For example, silicon has a diamond crystal structure with silicon atoms as shown in Figure 1.15a. Not only are three dimensions required, but also there is more than one atom per unit cell.

In addition charges associated with individual atoms in compound semiconductors depend on the degree of ionic character in the bonding. This will affect the detailed shape of the periodic potential. Also effects of electron shielding have not been accurately modelled. There are also other influences from electron spin and orbital angular momentum that influence energy bands in real crystals.

E versus *k* diagrams for various directions in a semiconductor crystal are often presented since the one-dimensional periodic potentials vary with direction. Although three-dimensional modelling is beyond the scope of this book the results for cubic crystals of silicon, germanium, gallium arsenide, gallium phosphide, gallium nitride and cadmium telluride as well as for wurtzite GaN are shown in Figure 1.16a–g. For cubic crystals these figures show the band shape for an electron travelling in the [111] crystal direction on the left side and for the [100] direction on the right side. It is clear that the periodic potential experienced by an electron travelling in various directions changes: the value of *a* in $u_k(x) = u_k(x + a)$ appropriate for use in the Bloch function (Equation 1.5) for the [100] direction is the edge length of the cubic unit cell of the crystal. For the [111] direction *a* must be modified to be the distance between the relevant atomic planes normal to the body diagonal of the unit cell. For wurtzite crystals the two directions shown are the [0001] direction along the *c*-axis and the $\langle 1100 \rangle$ directions along the *a*-axes.

Note that there are multiple valence bands that overlap or almost overlap with each other rather than a single valence band. These are *sub-bands* for holes, which are due to spin–orbit interactions that modify the band state energies for electrons in the valence band. The sub-bands are approximately parabolic near their maxima. Because the curvatures of these sub-bands vary, they give rise to what are referred to as *heavy holes* and *light holes* with m^* as described by Equation 1.16. There are also *split-off* bands with energy maxima below the valence band edge.

1.12   Direct Gap and Indirect Gap Semiconductors

In Figure 1.16 the conduction bands generally exhibit two energy minima rather than one minimum. Each local minimum can be approximated by a parabola whose curvature will determine the effective mass of the relevant electrons.

Referring to Figure 1.16c, we can see that the bandgap of GaAs is 1.43 eV where the valence band maximum and conduction band minimum coincide at $k = 0$. This occurs because the overall minimum of the conduction band is positioned at the same value of
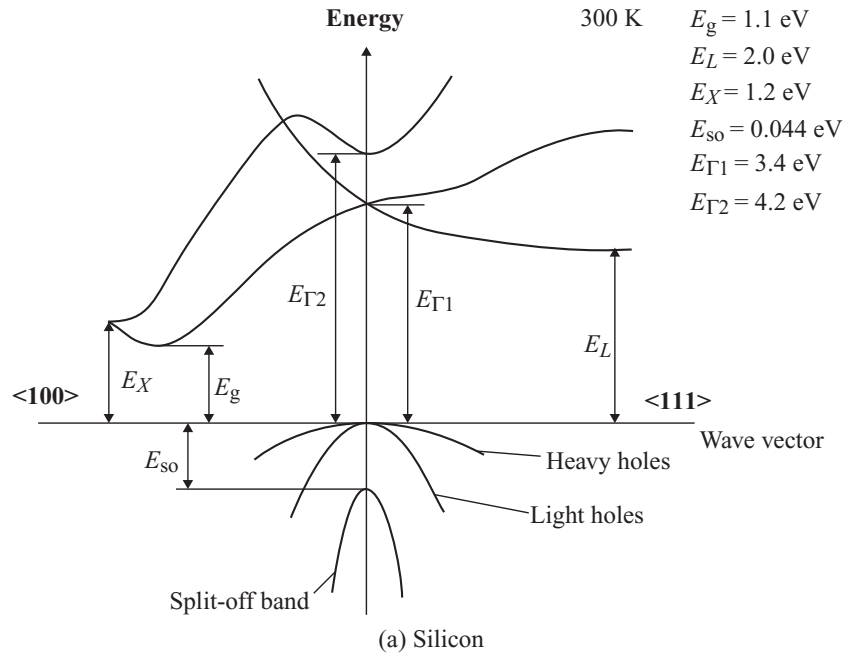
(a) Silicon

Figure 1.16   *Band structures of selected semiconductors. (a) silicon, (b) germanium, (c) GaAs, (d) GaP, (e) cubic GaN, (f) CdTe and (g) wurtzite GaN. Note that GaN is normally wurtzite. Cubic GaN is not an equilibrium phase at atmospheric pressure;however, it can be prepared at high pressure and it is stable once grown. Note that symbols are used to describe various band features. Γ denotes the point where $k = 0$. X and L denote the Brillouin zone boundaries in the $\langle 100 \rangle$ and $\langle 111 \rangle$ directions respectively in a cubic semiconductor. In (g) k_x and k_z denote the a and c directions respectively in a hexagonal semiconductor. See Figure 1.15c. Using the horizontal axes to depict two crystal directions saves drawing an additional figure; it is unnecessary to show the complete drawing for each k-direction since the positive and negative k-axes for a given k-direction are symmetrical. There are also energy gaps shown that are larger than the actual energy gap; the actual energy gap is the smallest gap. These band diagrams are the result of both measurements and modelling results. In some cases the energy gap values differ slightly from the values in Appendix 2. (a–d) Reprinted from Levinstein, M., Rumyantsev, S., and Shur, M., Handbook Series on Semiconductor Parameters vol. 1. ISBN 9810229348. Copyright (1996) with permission from World Scientific, London. (e, g) Reprinted from Morkoc, H., Handbook of Nitride Semiconductors and Devices, Vol. 1, ISBN 978-3-527-40837-5. Copyright (2008) WILEY-VCH Verlag GmbH & Co. KGaA Weinheim. (f) Reprinted from Chadov, S., et al., Tunable multifunctional topological insulators in ternary Heusler compounds, Nature Materials 9, 541–545. Copyright (2010) with permission from Nature Publishing Group. DOI: doi:10.1038/nmat2770*
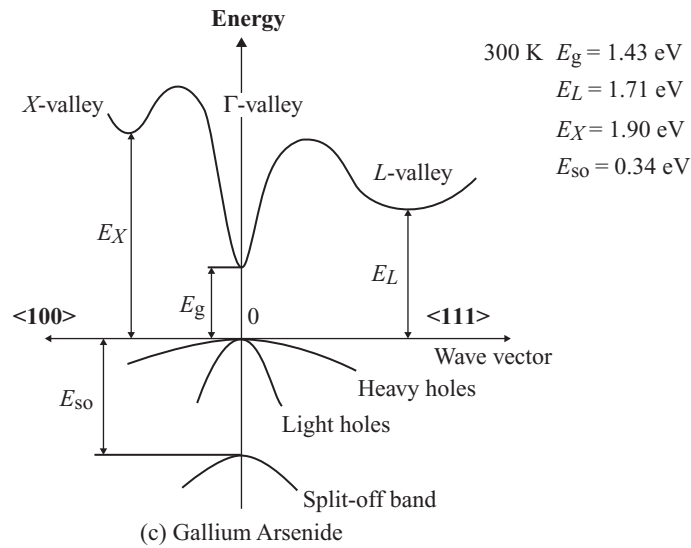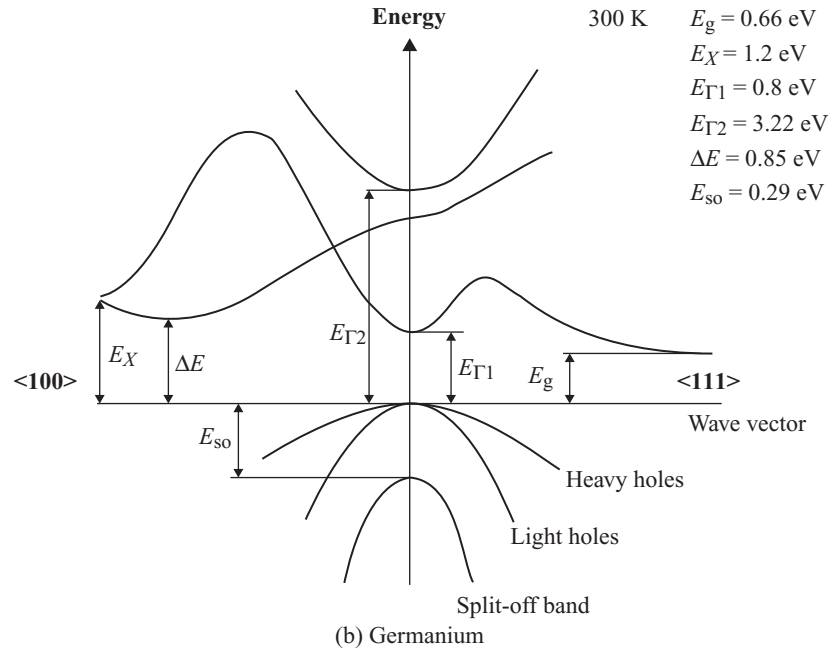
(b) Germanium

(c) Gallium Arsenide

Figure 1.16   *(Continued)*

Energy

300 K E_g = 2.26 eV

E_L = 2.6 eV

E_0 = 2.78 eV

E_{so} = 0.08 eV

Γ-valley Λ_{1C}

Γ_{1C} *L*-valley

X-valley L_{6C}

X_{1C}

E_0

E_1 E_L

<100> E_g E_2 Γ_{15V} **<111>**

E_{so} Wave vector

X_{5V} Λ_{3V}

Light holes   Heavy holes

Split-off band

(d) Gallium Phosphide

Zincblende   **Energy**

X-valleys   *L*-valleys

300 K
E_g = 3.2 eV
E_X = 4.6 eV
E_L = 4.8 - 5.1 eV
E_{so} = 0.02 eV

E_X Γ-valley E_L

E_g

0

<100> E_{so} Heavy holes   **<111>**

Light holes

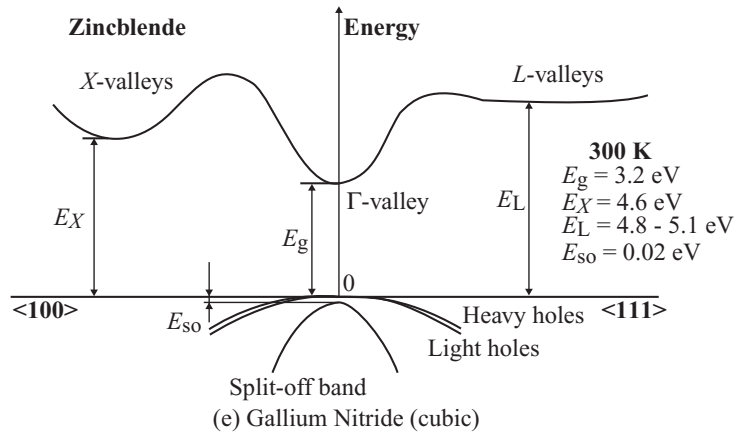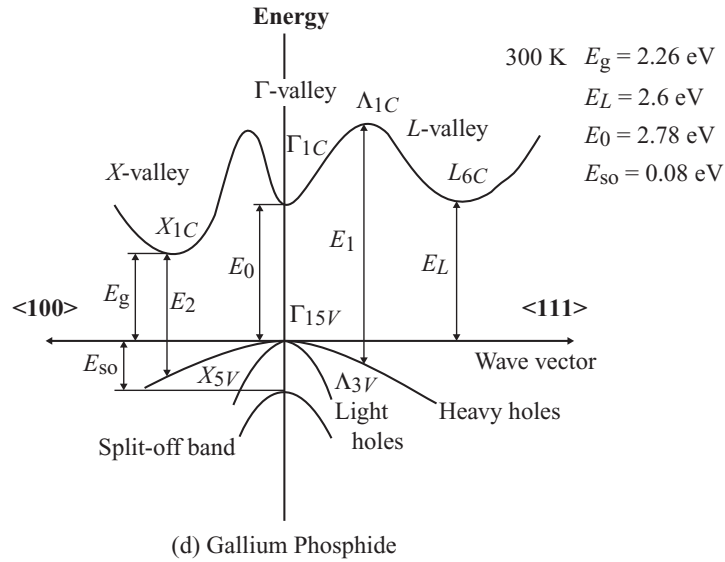Split-off band

(e) Gallium Nitride (cubic)

Figure 1.16   *(Continued)*

k as the valence band maximum and this results in a *direct gap* semiconductor. In Figure 1.16 GaAs, GaN and CdTe are direct gap semiconductors. In contrast to GaAs, silicon in Figure 1.16a has a valence band maximum at a different value of *k* than the conduction band minimum. That means that the energy gap of 1.1 eV is not determined by the separation between bands at $k = 0$, but rather by the distance between the overall conduction band minimum and valence band maximum. This results in an *indirect gap* semiconductor. Another indirect gap semiconductor in Figure 1.16 is the III-V material GaP.
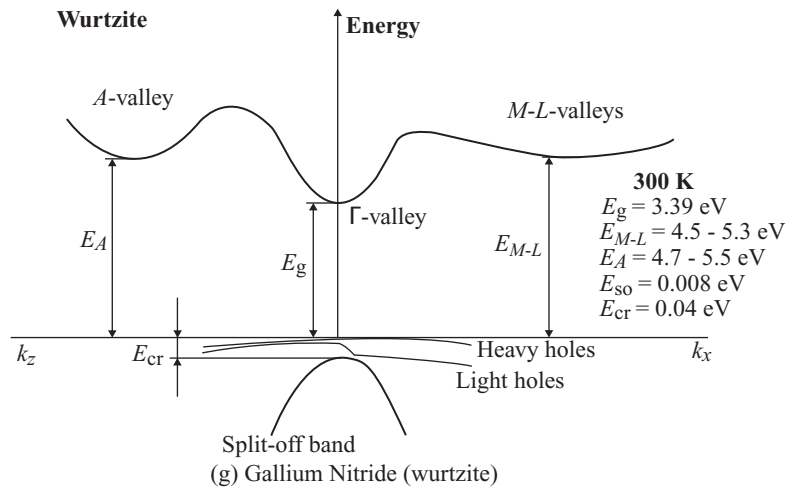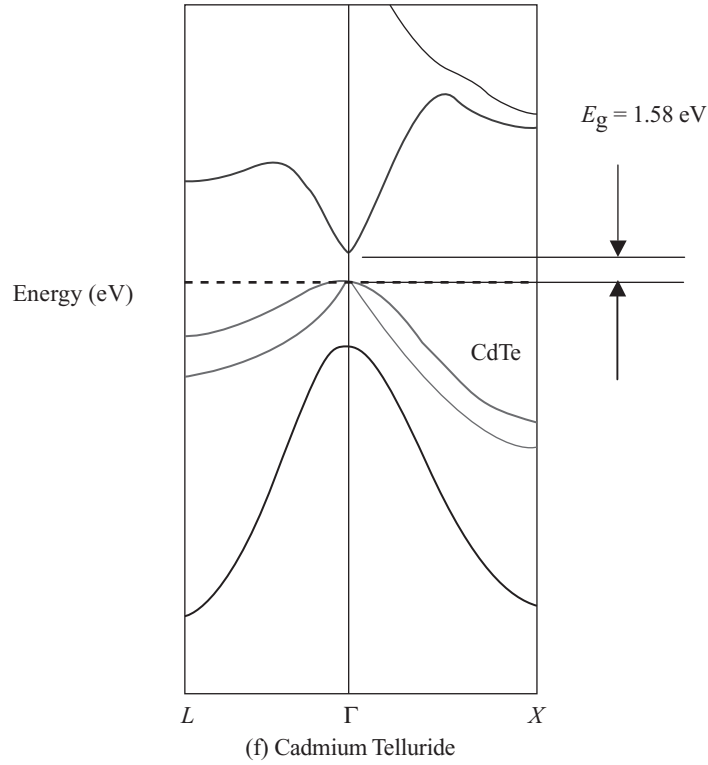
(f) Cadmium Telluride

(g) Gallium Nitride (wurtzite)

Figure 1.16   *(Continued)*

The distinction between direct and indirect gap semiconductors is of particular significance for photovoltaic and LED devices because processes involving photons occur in both cases, and photon absorption and generation properties differ considerably between these two semiconductor types.

An *electron-hole pair* (EHP) may be created if a photon is absorbed by a semiconductor and causes an electron in the valence band to be excited into the conduction band. For example, photon absorption in silicon can occur if the photon energy matches or exceeds the bandgap energy of 1.11 eV. Since silicon is an indirect gap semiconductor, however, there is a shift along the k axis for the electron that leaves the top of the valence band and then occupies the bottom of the conduction band. In Section 1.3 we noted that $p = \hbar k$ and therefore a shift in momentum results. The shift is considerable as seen in Figure 1.16a, and it is almost the distance from the centre of the Brillouin zone at $k = 0$ to the zone boundary at $k = \frac{\pi}{a}$ yielding a momentum shift of

$$\Delta p \simeq \hbar \frac{\pi}{a} \qquad (1.38)$$

During the creation of an EHP both energy and momentum must be conserved. Energy is conserved since the photon energy $\hbar\omega$ satisfies the condition $\hbar\omega = E_{\mathrm{g}}$. Photon momentum $p = \frac{h}{\lambda}$ is very small, however, and is unable to provide momentum conservation. This is discussed further in Section 4.2. This means that a lattice vibration, or *phonon*, is required to take part in the EHP generation process. The magnitudes of phonon momenta cover a wide range in crystals and a phonon with the required momentum may not be available to the EHP process. This limits the rate of EHP generation, and photons that are not absorbed continue to propagate through the silicon.

If electromagnetic radiation propagates through a semiconductor we quantify absorption using an *absorption coefficient* α, which determines the intensity of radiation by the exponential relationship

$$\frac{I(x)}{I_0} = \mathrm{e}^{-\alpha x}$$

where I_0 is the initial radiation intensity and $I(x)$ is the intensity after propagating through the semiconductor over a distance x. Efficient crystalline silicon solar cells are generally at least $\simeq 100$ μm thick for this reason due to their relatively low absorption coefficient. In contrast, GaAs (Figure 1.16c), is a direct gap semiconductor and has a much higher value of α (see Section 4.2). The thickness of GaAs required for sunlight absorption is only $\simeq 1$ μm. The value of α is an important parameter in PV semiconductors since sunlight that is not absorbed will not contribute to electric power generation. It is interesting to note that in spite of this difficulty silicon has historically been the most important solar cell material owing to its large cost advantage over GaAs.

In LEDs the process is reversed. EHPs recombine and give rise to photons, which are emitted as radiation. The wavelength range of this radiation may be in the infrared, the visible, or the ultraviolet parts of the electromagnetic spectrum, and is dependent on the semiconductor energy gap. Silicon is a poor material for LEDs because for an EHP recombination to create a photon, one or more phonons need to be involved to achieve momentum conservation. The probability for this to occur is therefore much smaller and competing mechanisms for electron-hole pair recombination become important. These are

known as *non-radiative recombination* events (see Section 1.19). In contrast to silicon, GaAs can be used for high-efficiency LEDs and was used for the first practical LED devices due to its direct gap.

1.13   Extrinsic Semiconductors

The incorporation of very small concentrations of impurities, referred to as *doping*, allows us to create semiconductors that are called *extrinsic* to distinguish them from intrinsic semiconductors, and we can control both the electron and hole concentrations over many orders of magnitude.

Consider the addition of a group V atom such as phosphorus to a silicon crystal as shown together with a band diagram in Figure 1.17. This results in an *n-type semiconductor*. The phosphorus atom substitutes for a silicon atom and is called a *donor*; it introduces a new spatially localized energy level called the donor level E_d.

Because phosphorus has one more electron than silicon this donor electron is not required for valence bonding, is only loosely bound to the phosphorus, and can easily be excited into the conduction band. The energy required for this is $E_c - E_d$ and is referred to as the *donor binding energy*. If the donor electron has entered the conduction band, it is no longer spatially localized and the donor becomes a positively charged ion. The donor binding energy may be calculated by considering the well-known hydrogen energy quantum states in which the ionization energy for a hydrogen atom is given by

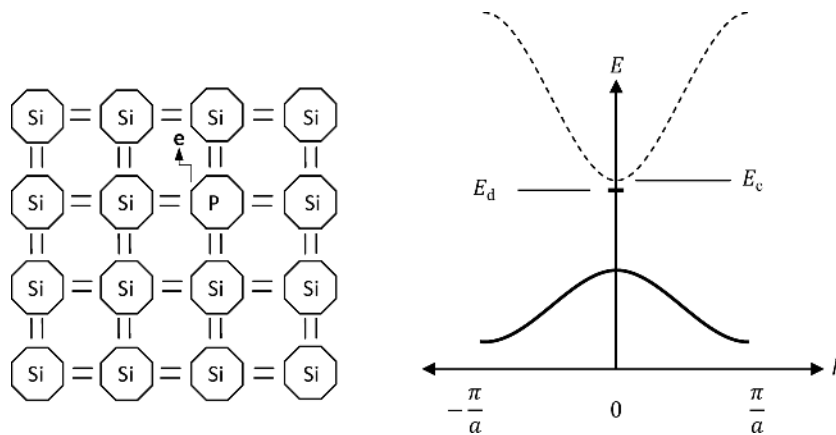$$E_{\text{Rydberg}} = \frac{-mq^4}{8\epsilon_0^2 h^2} = 13.6 \text{ eV} \tag{1.39}$$

Figure 1.17   *The substitution of a phosphorus atom in silicon (donor atom) results in a weakly bound extra electron occupying new energy level E_d that is not required to complete the covalent bonds in the crystal. It requires only a small energy $E_c - E_d$ to be excited into the conduction band, resulting in a positively charged donor ion and an extra electron in the conduction band*

and the Bohr radius given by

$$a_0 = \frac{4\pi\epsilon_0\,\hbar^2}{mq^2} = 0.529\ \text{Å} \tag{1.40}$$

Now, two variables in Equations 1.39 and 1.40 must be changed. Whereas the hydrogen electron moves in a vacuum, the donor is surrounded by semiconductor atoms, which requires us to modify the dielectric constant from the free space value ϵ_0 to the appropriate value for silicon by multiplying by the relative dielectric constant ϵ_r. In addition the free electron mass m must be changed to the effective mass m_e^*. This results in a small binding energy from Equation 1.39 compared to the hydrogen atom, and a large atomic radius from Equation 1.40 compared to the Bohr radius. For *n*-type dopants in silicon the measured values of binding energy are approximately 0.05 eV compared to 13.6 eV for the Rydberg constant, and an atomic radius is obtained that is an order of magnitude larger than the Bohr radius of approximately 0.5 Å. Since the atomic radius is now several lattice constants in diameter, we can justify the use of the bulk silicon constants we have used in place of vacuum constants.

Consider now the substitution of a group III atom such as aluminium for a silicon atom as illustrated in Figure 1.18. This creates a *p-type semiconductor*. The aluminium atom is called an *acceptor* and it introduces a new spatially localized energy level called the acceptor level E_a. Because aluminium has one fewer electron than silicon it can accept an electron from another valence bond elsewhere in the silicon, which results in a hole in the valence band. The energy required for this is $E_a - E_v$ and is referred to as the *acceptor binding energy*. If an electron has been accepted, the resulting hole is no longer spatially
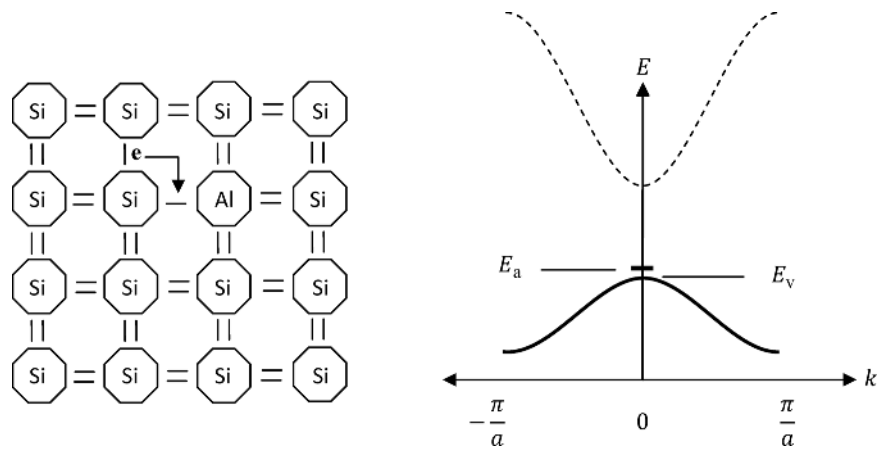

Figure 1.18   *The substitution of an aluminium atom in silicon (acceptor atom) results in an incomplete valence bond for the aluminium atom. An extra electron may be transferred to fill this bond from another valence bond in the crystal. The spatially localized energy level now occupied by this extra electron at E_a is slightly higher in energy than the valence band. This transfer requires only a small energy $E_a - E_v$ and results in a negatively charged acceptor ion and an extra hole in the valence band*

localized and the acceptor becomes a negatively charged ion. The binding energy may be estimated in a manner analogous to donor binding energies.

The introduction of either donors or acceptors influences the concentrations of charge carriers, and we need to be able to calculate these concentrations. The position of the Fermi level changes when dopant atoms are added, and it is no longer true that $n = p$; however, the Fermi–Dirac function $F(E)$ still applies. A very useful expression becomes the product of electron and hole concentrations in a given semiconductor. For intrinsic material, we have calculated $n_i p_i$ and we obtained Equation 1.37a; however, Equations 1.34 and 1.35 are still valid and we can also conclude that

$$n_0 p_0 = n_i^2 = p_i^2 = N_c N_v \exp\left(\frac{-E_g}{kT}\right) \tag{1.41}$$

which is independent of E_f, and therefore is also applicable to extrinsic semiconductors. Here n_0 and p_0 refer to the equilibrium carrier concentrations in the doped semiconductor.

We now examine the *intermediate temperature condition* where the following apply:

(a) The ambient temperature is high enough to ionize virtually all the donors or acceptors.
(b) The concentration of the dopant is much higher than the intrinsic carrier concentration because the ambient temperature is not high enough to directly excite a large number of electron-hole pairs.

Under these circumstances, there are two cases. For donor doping in an n-type semiconductor we can conclude that

$$n_0 \simeq N_d \tag{1.42}$$

and combining Equations 1.41 and 1.42 we obtain

$$p_0 = \frac{n_i^2}{N_d} \tag{1.43}$$

where N_d is the donor concentration in donor atoms per unit volume of the semiconductor. For acceptor doping in a p-type semiconductor we have

$$p_0 \simeq N_a \tag{1.44}$$

and we obtain

$$n_0 = \frac{n_i^2}{N_a} \tag{1.45}$$

The Fermi energy levels will change upon doping, and may be calculated from Equations 1.34 and 1.35. In the case of n-type silicon the Fermi level will lie closer to the conduction band. In the case of p-type silicon the Fermi level will lie closer to the valence band (see Figure 1.19). In Example 1.3 we calculate some specific values of the Fermi energy position.

Consider n-type silicon at room temperature. The mobile electrons in the n-type silicon are called *majority carriers*, and the mobile holes are called *minority carriers*. We can also consider p-type silicon with mobile holes called majority carriers and mobile electrons called minority carriers.
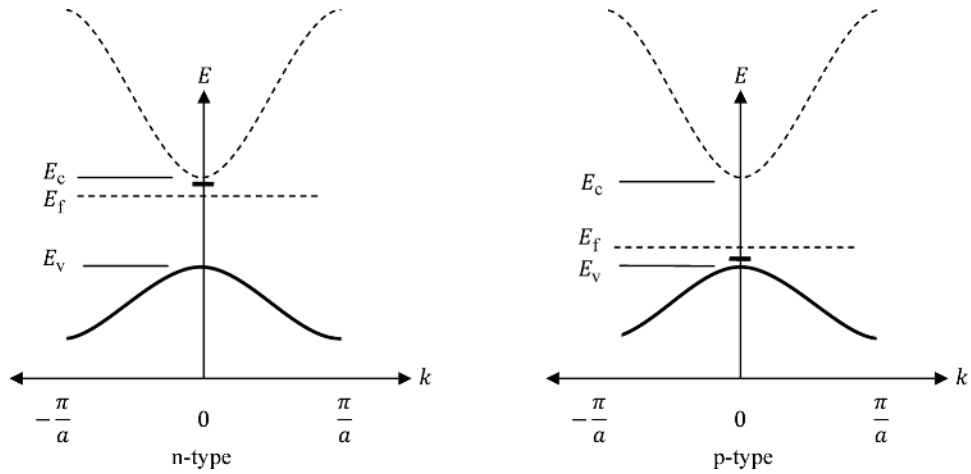
Figure 1.19   *The band diagrams for n-type silicon with a donor doping concentration of 1×10^{17} cm^{-3} and p-type silicon with an acceptor doping concentration of 1×10^{17} cm^{-3}. Note that the Fermi energy rises to the upper part of the energy gap for n-type doping and drops to the lower part of the energy gap for p-type doping*

Example 1.3

Assume a silicon sample at room temperature.

(a) Calculate the separation between E_c and E_f for n-type silicon having a phosphorus impurity concentration of 1×10^{17} cm^{-3}. Find both electron and hole concentrations.
(b) Calculate the separation between E_c and E_f for p-type silicon having an aluminium impurity concentration of 1×10^{17} cm^{-3}. Find both electron and hole concentrations.

Solution

(a) Using the intermediate temperature approximation $n_0 \cong 1 \times 10^{17}$ cm^{-3}. From example 1.2 $n_i = 1.5 \times 10^{10}$ cm^3 and hence
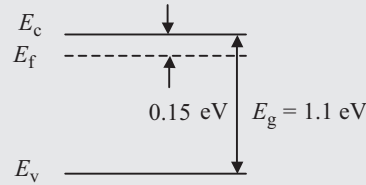
$$p_0 = \frac{n_i^2}{n_0} = \frac{(1.5 \times 10^{10} \text{ cm}^{-3})^2}{1 \times 10^{17} \text{ cm}^{-3}} = 2.25 \times 10^3 \text{ cm}^{-3}$$

and

$$n_0 = N_c \exp\left(\frac{-(E_c - E_f)}{kT}\right)$$

Solving for $E_c - E_f$,

$$E_c - E_f = kT \ln \frac{N_c}{n_0} = 0.026 \text{ eV} \ \ln\left(\frac{2.84 \times 10^{19} \text{ cm}^{-3}}{1 \times 10^{17} \text{ cm}^{-3}}\right) = 0.15 \text{ eV}$$

(b) Using the intermediate temperature approximation, $p_0 \cong 1 \times 10^{17} \text{ cm}^{-3}$ and hence
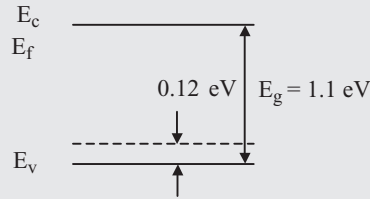
$$n_0 = \frac{n_i^2}{p_0} = \frac{(1.5 \times 10^{10} \text{ cm}^{-3})^2}{1 \times 10^{17} \text{ cm}^{-3}} = 2.25 \times 10^3 \text{ cm}^{-3}$$

and

$$p_0 = N_v \exp\left(\frac{-(E_f - E_v)}{kT}\right)$$

Solving for $E_f - E_v$

$$E_f - E_v = kT \ln \frac{N_v}{n_0} = 0.026 \text{ eV} \ \ln\left(\frac{1.06 \times 10^{19} \text{ cm}^{-3}}{1 \times 10^{17} \text{ cm}^{-3}}\right) = 0.12 \text{ eV}$$

At low ambient temperatures, the dopant atoms are not necessarily ionized. In this case there will be a combination of neutral and ionized dopant atoms. At high ambient temperatures, the intrinsic electron-hole pair concentration may be significant and may exceed the doping concentration. In this case the semiconductor carrier concentrations can be similar to intrinsic material. These cases are illustrated in Figure 1.20. Of particular technological importance is the intermediate temperature region since the carrier concentrations are relatively independent of temperature and therefore semiconductor devices can operate over wide temperature ranges without significant variation in carrier concentrations.
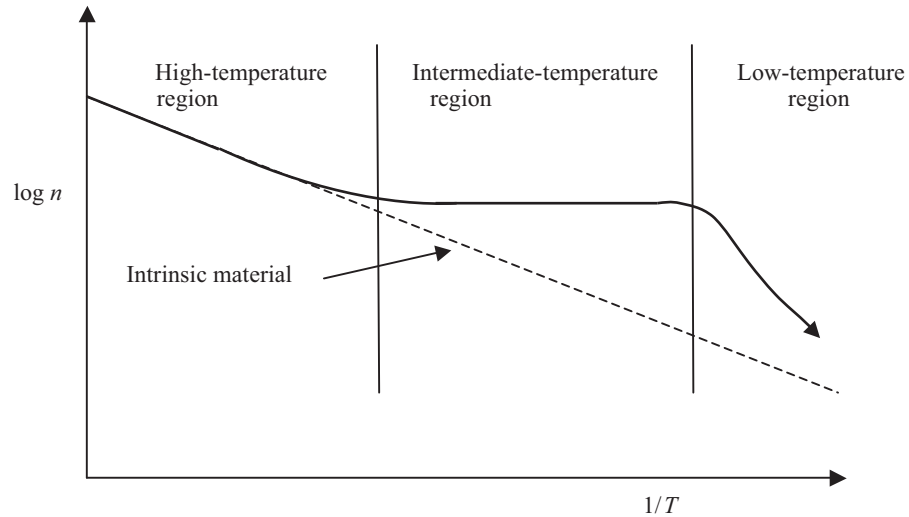
Figure 1.20 *Carrier concentration as a function of temperature for an n-type extrinsic semiconductor. In the high-temperature region behaviour is intrinsic-like. In the intermediate-temperature region carrier concentration is controlled by the impurity concentration and virtually all the dopant atoms are ionized. At low temperatures there is not enough thermal energy to completely ionize the dopant atoms*

1.14   Carrier Transport in Semiconductors

The electrical conductivity of semiconductors is controlled by the concentrations of both holes and electrons as well as their ability to flow in a specific direction under the influence of an electric field. The flow of carriers is limited by scattering events in which carriers having a high instantaneous velocity frequently scatter off lattice vibrations (phonons), defects and impurities, and we can denote a *scattering time* or characteristic mean time between scattering events for this, referred to as τ. The resulting net flow velocity or *drift velocity* of a stream of carriers is much lower than their instantaneous velocity. The experimental evidence for this is summarized by Ohm's law, or

$$J = \sigma \varepsilon$$

which is a collision-limited flow equation that relates the current flow to the applied electric field. To understand this we consider Figure 1.21 showing the flow of carriers in a solid cylinder of cross-sectional area A in the x direction.

If the carrier concentration is n and each carrier carrying charge q moves a distance dx in time dt then the amount of charge dQ passing across a given plane in the cylinder in time dt is $dQ = nqA dx$. The carrier drift velocity is given by $\bar{v} = \frac{dx}{dt}$, and we can conclude that the current is

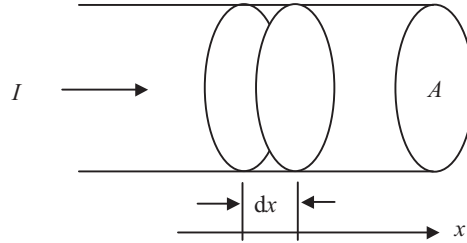$$I = \frac{dQ}{dt} = \frac{nqA dx}{dt} = nqA\bar{v}$$

Figure 1.21   *Current* (I) *flows along a solid semiconductor rod of cross-sectional area* A

We also define the current density $J = \frac{I}{A}$ and hence we obtain Ohm's law or

$$J = nq\bar{v} = \sigma\varepsilon$$

This is known as the *drift current* and it requires the existence of an electric field.

The application of an electric field can also be viewed using energy band diagrams. The well-known electrostatic relationship between electric field and electric potential and energy is given by

$$\varepsilon(x) = -\frac{\mathrm{d}V}{\mathrm{d}x} = \frac{1}{q}\frac{\mathrm{d}E}{\mathrm{d}x} \tag{1.46}$$

which states that an electric field causes a gradient in electric potential V and in addition an electric field causes a gradient in the potential energy E of a charged particle having charge q.

We can represent the conduction and valence bands in an applied electric field by showing the situation where the conduction and valence bands are separated by the energy gap. In Figure 1.7 this occurs at $k = 0$. We introduce spatial dependence by using the x-axis to show the position in the x direction of the semiconductor as in Figure 1.22. If there is no applied field the bands are simply horizontal lines. If a constant electric field is present the energy bands must tilt since from Equation 1.46 there will be a constant gradient in energy and the carriers in each band will experience a force F of magnitude $q\varepsilon$ in the directions shown and will travel so as to lower their potential energies. *The Fermi energy does not tilt*, since the electric field does not change the thermodynamic equilibrium.

We can now describe the flow of electrons. Since $\bar{v} \propto \varepsilon$ we write $\bar{v} = \mu\varepsilon$ where μ is the carrier *mobility*, and we also conclude from Ohm's law that $\sigma = nq\mu$.

In order to confirm the validity of Ohm's law we can start with Newton's law of motion for an electron in an electric field

$$\frac{F}{m^*} = -\frac{q\varepsilon}{m^*} = \frac{\mathrm{d}\bar{v}}{\mathrm{d}t}$$

The treatment of carrier collisions requires adding the well-known *damping term* $\frac{\bar{v}}{\tau}$ where τ is the scattering time that results in a terminal velocity. This can be pictured by the example of a terminal velocity reached by a ping-pong ball falling in air. We now have

$$\frac{F}{m^*} = -\frac{q\varepsilon}{m^*} = \frac{\mathrm{d}\bar{v}}{\mathrm{d}t} + \frac{\bar{v}}{\tau} \tag{1.47}$$
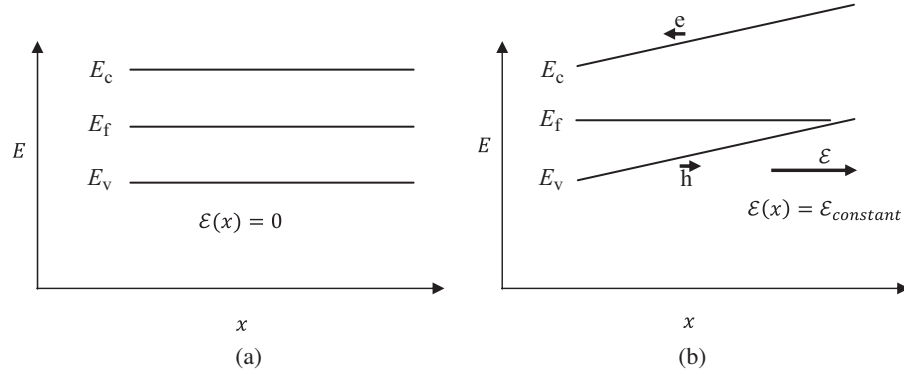
Figure 1.22 *Spatial dependence of energy bands in an intrinsic semiconductor. If there is no electric field (a) the bands are horizontal and electron and hole energies are independent of location within the semiconductor. If an electric field ε is present inside the semiconductor the bands tilt. For an electric field pointing to the right (b), electrons in the conduction band experience a force to the left, which decreases their potential energy. Holes in the valence band experience a force to the right, which decreases their potential energy. This reversed direction for hole energies is described in Figure 1.10*

We can demonstrate the validity of the equation in steady state where $\frac{d\bar{v}}{dt} = 0$ and hence

$$\bar{v} = \frac{q\tau\varepsilon}{m^*}$$

and

$$\mu = \frac{q\tau}{m^*}$$

which are consistent with Ohm's law.

In addition, we can examine the case where $\varepsilon = 0$. Now from Equation 1.47 we obtain

$$\frac{d\bar{v}}{dt} + \frac{\bar{v}}{\tau} = 0$$

which has solution $\bar{v}(t) = \bar{v}(0)\exp\left(-\frac{t}{\tau}\right)$. Carrier drift velocity will decay upon removal of the electric field with characteristic time constant equal to the scattering time τ.

In order to consider the contribution of both electrons and holes, we write the total drift current as

$$J_{\text{drift}} = J_{\text{n}} + J_{\text{p}} = q(n\mu_n + p\mu_p)\varepsilon \qquad (1.48)$$

where

$$\mu_{\text{n}} = \frac{q\tau}{m_{\text{n}}^*}$$

and

$$\mu_{\text{p}} = \frac{q\tau}{m_{\text{p}}^*}$$

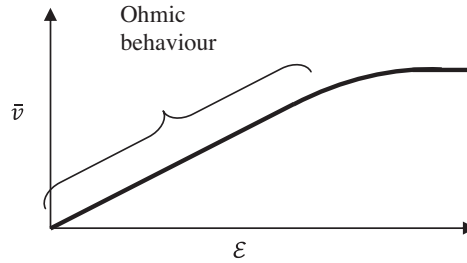**Figure 1.23**   *Dependence of drift velocity on electric field for a semiconductor*

Separate mobility values μ_n and μ_p are needed for electrons and holes since they flow in different bands and may have different effective masses m_n^* and m_p^* respectively. The valence band has negative curvature, and Equation 1.16 shows that valence band electrons have negative effective mass; however, to ensure that μ_h is a positive quantity we define the effective mass of holes m_h^* to be a *positive* quantity equal in magnitude to this negative effective mass.

The validity of Ohm's law has a limit. If the electric field is large, carrier velocity will usually saturate, as shown in Figure 1.23 and will no longer be linearly proportional to the electric field. This occurs because energetic electrons transfer more energy to lattice vibrations. The magnitude of the electric field that results in saturation effects depends on the semiconductor.

Mobility values for a range of semiconductors are shown in Appendix 2. These are tabulated for intrinsic materials at room temperature; however, defects and impurities as well as higher temperatures have a substantial effect on mobility values since they decrease scattering times. Both undesirable impurities as well as intentionally introduced dopant atoms will cause scattering times and mobility values to decrease.

1.15   Equilibrium and Non-Equilibrium Dynamics

The carrier concentrations we have been discussing until now are *equilibrium* concentrations, and are in thermodynamic equilibrium with the semiconductor material. In equilibrium, both EHP generation and EHP recombination occur simultaneously; however, the net EHP concentration remains constant. We can express this using rate constants defined as follows:

$$G_{th} = \text{thermal EHP generation rate (EHP cm}^{-3}\text{ s}^{-1})$$
$$R = \text{EHP recombination rate (EHP cm}^{-3}\text{ s}^{-1})$$

In equilibrium $G_{th} = R$.

It is easy to cause a *non-equilibrium* condition to exist in a semiconductor. For example, we can illuminate the semiconductor with photons whose energy exceeds its energy gap of the semiconductor. We can also cause electric current to flow through the semiconductor by attaching two or more electrodes to the semiconductor and then connecting them across a voltage source.

In an ideal direct gap semiconductor the value of R depends on carrier concentrations. If, for example, the electron concentration n is doubled, R will double since the probability for an electron to reach a hole has doubled. If both n and p are doubled then R will increase by a factor of four since the hole concentration has also doubled. We can state this mathematically as $R \propto np$. We remove subscripts on carrier concentrations when non-equilibrium conditions are present.

If a steady optical generation rate G_{op} is abruptly added to the thermal generation rate by illumination that begins at time $t = 0$, the total generation rate increases to $G_{th} + G_{op}$. This will cause the EHP generation rate to exceed the recombination rate and carrier concentrations will exceed the equilibrium concentrations and will become time dependent. We shall designate $\delta n(t)$ and $\delta p(t)$ to be the time-dependent carrier concentrations in excess of equilibrium concentrations n_0 and p_0.

Consider a direct gap n-type semiconductor in which $n_0 \gg p_0$. We shall also stipulate that $\delta n(t) \ll n_0$, which states that the excess carrier concentration is small compared to the equilibrium majority carrier concentration. This is referred to as the *low-level injection* condition. We can approximate n to be almost constant and essentially independent of the illumination. Hence the rate of increase of p is determined by the optical generation rate as well as a hole recombination rate term that is linearly proportional to $\delta p(t)$ and we can write the simple differential equation:

$$\frac{d\delta p(t)}{dt} = G_{op} - \frac{\delta p(t)}{\tau_p} \tag{1.49}$$

The solution to this is

$$\delta p(t) = \Delta p \left[1 - \exp\left(\frac{-t}{\tau_p} \right) \right] \tag{1.50}$$

which is sketched in Figure 1.24.

After a time $t \gg \tau_p$, a steady-state value of excess carrier concentration Δp exists. If Equation 1.50 is substituted into Equation 1.49 we obtain

$$\Delta p = G_{op} \tau_p \tag{1.51a}$$

The time constant τ_p for this process is called the *recombination time*. Recombination time is often the same as the *minority carrier lifetime* but should not be confused with the scattering time of Section 1.14, which is generally orders of magnitude shorter than the recombination time. From Equation 1.50, if we know the optical generation rate and
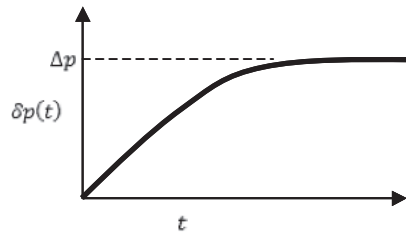

Figure 1.24   *Plot of excess hole concentration as a function of time. A constant optical generation rate starts at* t = 0 *and continues indefinitely*

the characteristic recombination time for a semiconductor, we can calculate the steady-state excess minority carrier concentration. There is also a small steady-state change in the majority carrier concentration because $\Delta n = \Delta p$; however, this is neglected in low-level injection since $\Delta n \ll n_0$. Note that $\tau_p \propto \frac{1}{n_0}$.

The same argument can be applied to a p-type semiconductor, and we would obtain the increase in minority carrier concentration thus:

$$\Delta n = G_{\text{op}} \tau_{\text{n}} \tag{1.51b}$$

Electron-hole recombination may also occur via indirect processes in indirect gap semiconductors such as silicon. This is discussed in section 1.19.

1.16   Carrier Diffusion and the Einstein Relation

Free carriers that are produced in a spatially localized part of a semiconductor are able to diffuse and thereby move to other parts of the material. The carrier diffusion process is functionally similar to the diffusion of atoms in solids. At sufficiently high temperatures atomic diffusion occurs, which is described as net atomic motion from a region of higher atomic concentration to a region of lower atomic concentration. This occurs due to random movements of atoms in a concentration gradient. There is no preferred direction to the random movement of the atoms; however, provided the average concentration of atoms is not uniform, the result of random movement is for a net flux of atoms to exist flowing from a more concentrated region to a less concentrated region. *Fick's first law* applies to the diffusion process for atoms as presented in introductory materials science textbooks, and is also applicable to electrons.

As with atomic diffusion, the driving force for carrier diffusion is the gradient in electron concentration. For free electrons diffusing along the *x*-axis, Fick's first law applies and it can be written

$$\phi_{\text{n}}(x) = -D_{\text{n}} \frac{dn(x)}{dx} \tag{1.52}$$

where ϕ_{n} is the flux of electrons (number of electrons per unit area per second) flowing along the *x*-axis due to a concentration gradient of electrons. The negative sign in Equation 1.52 indicates that diffusion occurs in the direction of decreasing electron concentration. For holes, Fick's first law becomes

$$\phi_{\text{p}}(x) = -D_{\text{p}} \frac{dp(x)}{dx} \tag{1.53}$$

Since the flow of charged particles constitutes an electric current, we can describe *diffusion currents* due to holes or electrons. These are distinct from drift currents described in Section 1.14 because no electric field is involved. Equations 1.52 and 1.53 may be rewritten as currents:

$$J_{\text{n}}(x)_{\text{diffusion}} = q D_{\text{n}} \frac{dn(x)}{dx} \tag{1.54a}$$

$$J_{\text{p}}(x)_{\text{diffusion}} = -q D_{\text{p}} \frac{dp(x)}{dx} \tag{1.54b}$$

Note that there is no negative sign in the case of electrons because electrons carry a negative charge that cancels out the negative sign in Fick's first law.

An interesting situation occurs when both diffusion and drift currents flow. An electric field is present as well as a carrier concentration gradient. The total current densities from Equations 1.48 and 1.54 become

$$J_n(x)_{\text{drift+diffusion}} = q\mu_n n(x)\varepsilon(x) + qD_n\frac{dn(x)}{dx} \qquad (1.55a)$$

and

$$J_p(x)_{\text{drift+diffusion}} = q\mu_p p(x)\varepsilon(x) - qD_p\frac{dp(x)}{dx} \qquad (1.55b)$$

In semiconductor diodes, both drift and diffusion occur and it is important to become familiar with the situation where drift and diffusion currents coexist in the same part of the semiconductor.

One way to establish an electric field is to have a gradient in doping level by spatially varying the doping concentration in the semiconductor. Consider the example in Figure 1.25. The dopant concentration varies across a semiconductor sample that is in thermal equilibrium. On the left side, the semiconductor is undoped, and an acceptor dopant gradually increases in concentration from left to right. This causes the Fermi energy to occupy lower positions in the energy gap until it is close to the valence band on the right
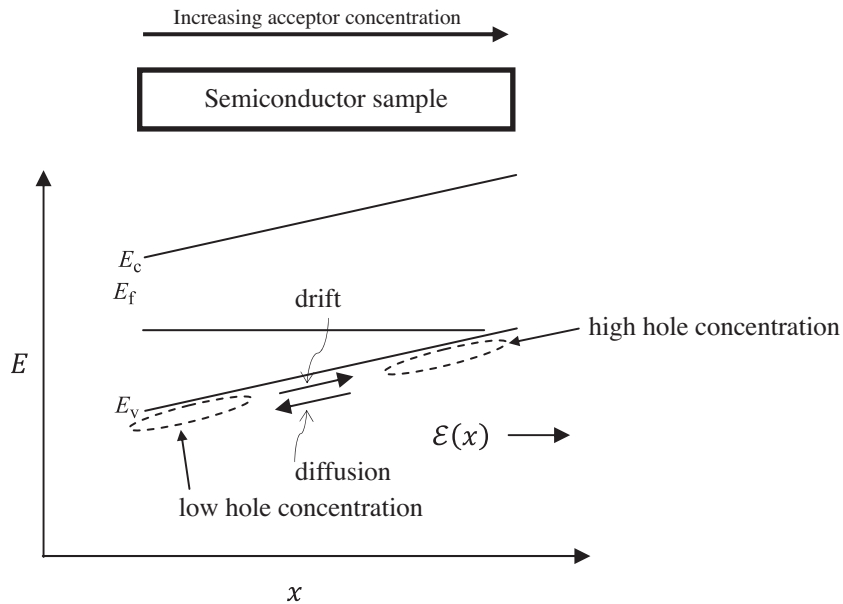

Figure 1.25 *The energy bands will tilt due to a doping gradient. Acceptor concentration increases from left to right in a semiconductor sample. This causes a built-in electric field, and the hole concentration increases from left to right. The field causes hole* drift *from left to right, and there is also hole* diffusion *from right to left due to the concentration gradient*

side of the sample. The Fermi energy does not tilt because it is a thermodynamic quantity and the sample is in equilibrium; however, the valence and conduction bands do tilt as shown.

There will now be a high concentration of holes in the valence band on the right side of the semiconductor, which decreases to a low hole concentration on the left side. Hole diffusion will therefore occur in the negative-*x* direction. At the same time, the tilting of the energy bands means that an electric field is present in the sample. This is known as a *built-in electric field* since it is caused by a spatial concentration variation within the semiconductor material rather than by the application of an applied voltage. The built-in field causes a hole drift current to flow in the positive-*x* direction. Since the semiconductor is in equilibrium, these two hole currents cancel out and *the net hole current flow will be zero*. A similar argument can be made for the electrons in the conduction band and the net electron current will also be zero.

A useful relationship between mobility and diffusivity can now be derived. Since the net current flow illustrated in Figure 1.25 is zero in equilibrium we can write for hole current

$$J_p(x)_{\text{drift+diffusion}} = q\mu_p p(x)\varepsilon(x) - qD_p\frac{dp(x)}{dx} = 0 \qquad (1.56)$$

From Equations 1.46 and 1.56 we have

$$q\mu_p p(x)\frac{1}{q}\frac{dE}{dx} - qD_p\frac{dp(x)}{dx} = 0$$

We now calculate $p(x)$. Since the valence band energy E_v is now a function of x we can rewrite Equation 1.35 as

$$p(x) = N_v \exp\left(\frac{-(E_f - E_v(x))}{kT}\right) \qquad (1.57)$$

and we obtain

$$\mu_p N_v \exp\left(\frac{-(E_f - E_v(x))}{kT}\right)\frac{dE_v(x)}{dx} - qD_p\frac{N_v}{kT}\exp\left(\frac{-(E_f - E_v(x))}{kT}\right)\frac{dE_v(x)}{dx} = 0$$

which simplifies to

$$\frac{D_p}{\mu_p} = \frac{kT}{q} \qquad (1.58a)$$

and a similar derivation may be applied to electrons yielding

$$\frac{D_n}{\mu_n} = \frac{kT}{q} \qquad (1.58b)$$

Equation 1.58 is known as the *Einstein relation*. At a given temperature this tells us that mobility and diffusivity are related by a constant factor, which is not unexpected since both

quantities express the degree of ease with which carriers move in a semiconductor under a driving force.

1.17 Quasi-Fermi Energies

If a semiconductor is influenced by incident photons or an applied electric current the semiconductor is no longer in equilibrium. This means that we cannot use Equations 1.34 and 1.35 to determine carrier concentrations. In addition Fermi energy E_f is no longer a meaningful quantity since it was defined for a semiconductor in equilibrium in Section 1.8 and the Fermi–Dirac distribution function of Figure 1.9 is also based on equilibrium conditions.

For convenience, we define two new quantities, F_n and F_p, known as the *quasi-Fermi energy for electrons* and the *quasi-Fermi energy for holes*, respectively. The quantities may be used even if a semiconductor is not in equilibrium and there are excess carriers. F_n and F_p are defined for a semiconductor with excess carriers from the following equations:

$$n = N_c \exp\left(\frac{-(E_c - F_n)}{kT}\right) \tag{1.59a}$$

and

$$p = N_v \exp\left(\frac{-(F_p - E_v)}{kT}\right) \tag{1.59b}$$

Note the similarity between Equation 1.59 and Equations 1.34 and 1.35. Subscripts for n and p are absent in Equation 1.59 because these carrier concentrations are not necessarily equilibrium values. It follows that if the semiconductor is in equilibrium, the electron and hole quasi-Fermi energies become equal to each other and identical to the Fermi energy. The electron and hole quasi-Fermi levels in an n-type semiconductor will behave very differently upon excess carrier generation, as shown in Figure 1.26. This is examined in Example 1.4.
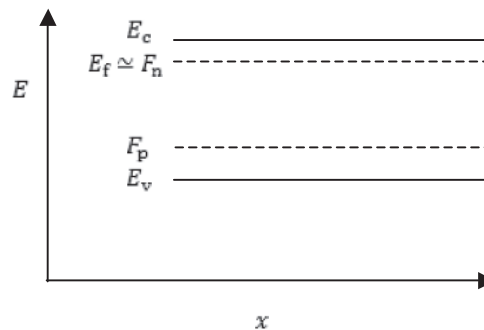

Figure 1.26   *The quasi-Fermi levels* F_n *and* F_p *for an n-type semiconductor with excess carriers generated by illumination. Note the large change in* F_p *due to illumination and note that* F_n *is almost the same as the value of* E_f *before illumination*

Example 1.4

An n-type silicon sample has a donor concentration of 1×10^{17} cm^{-3}. In Example 1.3a we obtained $n_0 = 1 \times 10^{17}$ cm^{-3}, $p_0 = 2.25 \times 10^3$ cm^{-3} and $E_c - E_f = 0.15$ eV. We now illuminate this sample and introduce a uniform electron-hole pair generation rate of $G_{op} = 5 \times 10^{20}$ cm^{-3} s^{-1}. Assume a carrier lifetime of 2×10^{-6} s.

(a) Calculate the resulting electron and hole concentrations.
(b) Calculate the quasi-Fermi energy levels.

Solution

(a)

$$\Delta p = G_{op}\tau_p = 5 \times 10^{20} \text{ cm}^{-3} \text{ s}^{-1} \times 2 \times 10^{-6} \text{ s} = 1 \times 10^{15} \text{ cm}^{-3}$$

and

$$\Delta n = G_{op}\tau_n = 5 \times 10^{20} \text{ cm}^{-3} \text{ s}^{-1} \times 2 \times 10^{-6} \text{ s} = 1 \times 10^{15} \text{ cm}^{-3}$$

Hence

$$p = p_0 + \Delta p = 2.25 \times 10^3 \text{ cm}^{-3} + 1 \times 10^{15} \text{ cm}^{-3} \cong 1 \times 10^{15} \text{ cm}^{-3}$$

and

$$n = n_0 + \Delta n = 1 \times 10^{17} \text{ cm}^{-3} + 1 \times 10^{15} \text{ cm}^{-3} = 1.01 \times 10^{17} \text{ cm}^{-3}$$

Therefore the carrier concentrations may be strongly affected by the illumination: the hole concentration increases by approximately 12 orders of magnitude from a very small minority carrier concentration to a much larger value dominated by the excess hole concentration.

The electron concentration, however, only increases slightly (by 1%) due to the illumination since it is a majority carrier. This is therefore an example of low-level injection since the majority carrier concentration is almost unchanged.

(b) The quasi-Fermi level for holes may be found from:

$$p = N_v \exp\left(\frac{-(F_p - E_v)}{kT}\right)$$
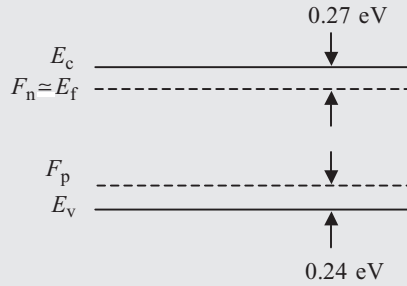
Solving for $F_p - E_v$ we obtain

$$F_p - E_v = kT \ln\left(\frac{N_v}{p}\right) = 0.026 \text{ eV } \ln\left(\frac{1.06 \times 10^{19} \text{ cm}^{-3}}{1 \times 10^{15} \text{ cm}^{-3}}\right) = 0.24 \text{ eV}$$

The quasi-Fermi level for electrons may be found from

$$n = N_c \exp\left(\frac{-(E_c - F_n)}{kT}\right)$$

Solving for $E_c - F_n$ we obtain

$$E_c - F_n = kT \ln\left(\frac{N_c}{n}\right) = 0.026 \text{ eV} \ \ln\left(\frac{2.84 \times 10^{19} \text{ cm}^{-3}}{1.01 \times 10^{15} \text{ cm}^{-3}}\right) = 0.27 \text{ eV}$$

0.27 eV

E_c ——————————————

$F_n \simeq E_f$ -

F_p -
E_v ——————————————

0.24 eV

Note that with illumination F_n is almost identical to the original value of E_f but F_p moves significantly lower. This is a consequence of the large excess carrier concentration compared to the equilibrium hole concentration. If a similar problem were solved for a p-type semiconductor then with illumination F_n would move significantly higher and F_p would remain almost identical to the original value of E_f.

The separation between F_n and F_p is a measure of the product of excess carrier concentrations divided by the product of equilibrium carrier concentrations since from Equations 1.59 and 1.41 we have

$$np = N_c N_v \exp\left(\frac{F_n - F_p - E_g}{kT}\right) = n_0 p_0 \exp\left(\frac{F_n - F_p}{kT}\right) \tag{1.60a}$$

and hence

$$F_n - F_p = kT \ln\left(\frac{np}{n_0 p_0}\right) \tag{1.60b}$$

1.18   The Diffusion Equation

We have introduced carrier recombination as well as carrier diffusion separately; however, carriers in semiconductors routinely undergo both diffusion and recombination simultaneously.

In order to describe this, consider a long semiconductor bar or rod in Figure 1.27 in which excess holes are generated at $x = 0$ causing an excess of holes Δp to be maintained at $x = 0$. The excess hole concentration drops off to approach an equilibrium concentration at the other end of the rod. The excess holes will diffuse to the right and if an electric field is present there will also be a drift current component. Some of these holes recombine
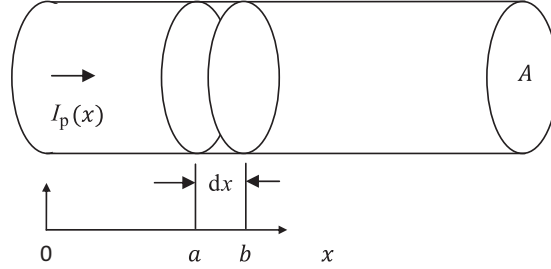
Figure 1.27 *A solid semiconductor rod of cross-sectional area* A *has a hole current* I$_p$(x) *flowing in the positive* x *direction. Due to recombination the hole current is dependent on* x. *At surfaces* x = a *and* x = b, I$_p$(x) *changes due to the recombination that occurs between these of the rod surfaces within volume* Adx

with electrons during this process. We can consider a slice of width dx as shown in Figure 1.27. The hole current $I_p (x = a)$ will be higher than the hole current $I_p (x = b)$ due to the rate of recombination of holes in volume Adx between $x = a$ and $x = b$. Volume Adx contains $Adx\delta p(x)$ excess holes. Since $\frac{\delta p(x)}{\tau_p}$ is the recombination rate this may be expressed mathematically:

$$I_p (x = a) - I_p (x = b) = -qAdx\frac{\delta p(x)}{\tau_p} \tag{1.61}$$

This may be rewritten as

$$\frac{I_p (x = a) - I_p (x = b)}{dx} = -qA\frac{\delta p(x)}{\tau_p}$$

or

$$\frac{dI_p(x)}{dx} = -qA\frac{\delta p(x)}{\tau_p}$$

In terms of current density we have

$$\frac{dJ_p(x)}{dx} = -q\frac{\delta p(x)}{\tau_p} \tag{1.62a}$$

and applying the same procedure to electrons we obtain

$$\frac{dJ_n(x)}{dx} = -q\frac{\delta n(x)}{\tau_n} \tag{1.62b}$$

If the current is entirely due to the *diffusion* of carriers, we rewrite the expression for diffusion current from Equation 1.54 for excess carriers $\delta p(x)$:

$$J_p(x)_{\text{diffusion}} = -qD_p\frac{d\delta p(x)}{dx} \tag{1.63}$$

Substituting this into Equation 1.62a we obtain

$$\frac{d^2 \delta p(x)}{dx^2} = \frac{\delta p(x)}{D_p \tau_p} \tag{1.64a}$$

This is known as the *steady state diffusion equation for holes*, and the corresponding equation for electrons is:

$$\frac{d^2 \delta n(x)}{dx^2} = \frac{\delta n(x)}{D_n \tau_n} \tag{1.64b}$$

The general solution to Equation 1.64a is

$$\delta p(x) = A \exp\left(\frac{-x}{\sqrt{D_p \tau_p}}\right) + B \exp\left(\frac{x}{\sqrt{D_p \tau_p}}\right) \tag{1.65a}$$

However, considering our boundary conditions, the function must decay to zero for large values of x and therefore $B = 0$ yielding

$$\delta p(x) = \Delta p \exp\left(\frac{-x}{\sqrt{D_p \tau_p}}\right)$$

which may be written

$$\delta p(x) = \Delta p \exp\left(\frac{-x}{L_p}\right) \tag{1.65b}$$

where

$$L_p = \sqrt{D_p \tau_p}$$

is known as the *diffusion length*. The latter determines the position on the x-axis where carrier concentrations are reduced by a factor of e, as shown in Figure 1.28.
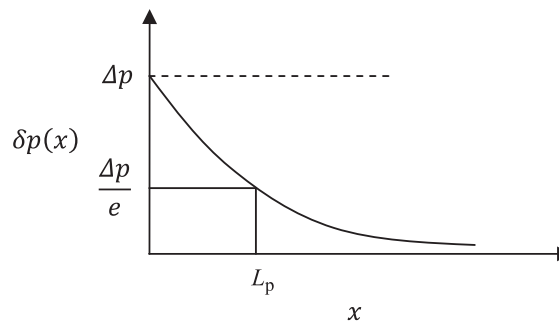

Figure 1.28 *Plot of excess hole concentration in a semiconductor as a function of x in a semiconductor rod where both diffusion and recombination occur simultaneously. The decay of the concentration is characterized by a diffusion length L_p*
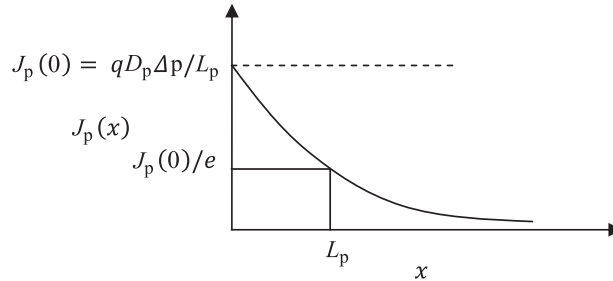
Figure 1.29 *Hole current density as a function of* x *for a semiconductor rod with excess carriers generated at* x = 0

The hole current density at any point x may be determined by substituting Equation 1.65b into Equation 1.63, and we obtain

$$J_\text{p}(x)_\text{diffusion} = q\frac{D_\text{p}}{L_\text{p}}\Delta p \, \exp\left(\frac{-x}{L_\text{p}}\right)$$

This may also be written

$$J_\text{p}(x)_\text{diffusion} = q\frac{D_\text{p}}{L_\text{p}}\delta p(x)$$

which shows that both current density and hole concentration have the same exponential form. Figure 1.29 plots current density $J_\text{p}(x)_\text{diffusion}$ as a function of x. Both diffusion and recombination occur simultaneously, which lowers the diffusion current exponentially as x increases. This is reasonable since the number of holes that have not recombined drops exponentially with x and therefore L_p also represents the position on the x-axis where current density $J_\text{p}(x)$ is reduced by a factor of e.

1.19   Traps and Carrier Lifetimes

Carrier lifetimes in many semiconductors including indirect gap silicon are really determined by trapping processes instead of ideal direct gap electron-hole recombination. *Traps* are impurity atoms or native point defects such as vacancies, dislocations or grain boundaries. There are also *surface traps* due to the defects that inherently occur at semiconductor surfaces, and *interface traps* that form at a boundary between two different material regions in a semiconductor device.

In all cases, traps are physical defects capable of trapping conduction band electrons and/or valence band holes and affecting carrier concentrations, carrier flow and recombination times. This is different from a normal electron-hole pair recombination process because a specific defect is involved. After being trapped a carrier may again be released to the band it originated from or it may subsequently recombine at the trap with a carrier of the opposite sign that also gets attracted to the same trap. This is known as *trap-assisted carrier recombination* and is one of the most important phenomena that limits the performance of both direct and indirect gap semiconductors used for solar cells and LEDs.

Intentionally introduced n-type and p-type dopants actually are traps, and are referred to as *shallow traps* because they are only separated from either a conduction band or a valence band by a small energy difference that may be overcome by thermal energy. This means that the trap is easy to ionize and the carrier is very likely to be released from the trap. Also since dopants are normally ionized, they do not trap carriers of the opposite sign. For example, phosphorus in silicon is an n-type dopant. Once it becomes a positive ion after donating an electron to the conduction band it has a small but finite probability of recapturing a conduction band electron. Since it spends almost all its time as a positive ion, it has essentially no chance of capturing a hole, which is repelled by the positive charge, and therefore shallow traps generally do not cause trap-assisted carrier recombination.

The traps that we must pay careful attention to are *deep traps*, which exist near the middle of the bandgap. These traps are highly effective at promoting electron-hole pair recombination events. Since their energy levels are well separated from band edges, carriers that are trapped are not easily released. Imagine a deep trap that captures a conduction band electron and is then negatively charged. In this state the negatively charged trap cannot readily release its trapped electron and may therefore attract a positive charge and act as an effective hole trap. Once the hole is trapped it recombines with the trapped electron and the trap is effectively emptied and is again available to trap another conduction band electron. In this manner, traps become a new conduit for electrons and holes to recombine. If the deep trap density is high the average trap-assisted recombination rate is high.

The analysis of deep trap behaviour requires that we know the probabilities of the trap being filled or empty. This may be understood in equilibrium conditions by knowing the trap energy level and comparing it to the Fermi energy level. If the trap energy level E_t is above E_f then the trap is more likely to be empty than full. If E_t is below E_f it is more likely to be filled. The terms 'filled' and 'empty' refer specifically to electrons because the Fermi–Dirac function describes the probability that an electron fills a specific energy level. If we wish to describe the probabilities for a trap to be occupied by holes, we must subtract these probabilities from 1.

We will simplify the treatment of traps by focusing on a very specific situation. Consider a trap *at the Fermi energy and near mid-gap* in a semiconductor. Both E_t and E_f will be at approximately the middle of the energy gap. The ionization energy of the trap for either a trapped electron or a trapped hole is approximately $\frac{E_g}{2}$, as illustrated in Figure 1.30. The probability that the trap is empty or filled is 50% since it is at the Fermi energy. The captured electron may be re-released back to the conduction band, or it may be annihilated by a hole at the trap.

There is a simple argument for assuming the trap is likely to exist at the Fermi level and near mid-gap. At the surface of a semiconductor, approximately half the bonds normally formed will be broken or incomplete. This gives rise to electrons that are only held by half the atoms that they usually are associated with in the interior of a perfect crystal. These
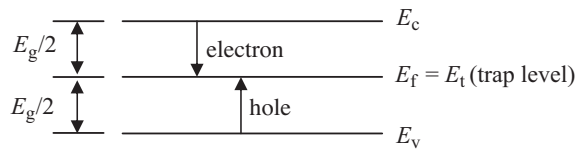

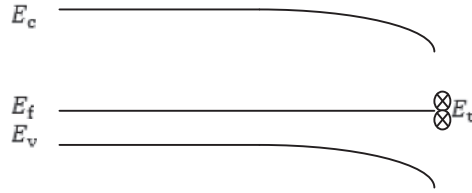**Figure 1.30**   *A trap level at the Fermi energy near mid-gap*

Figure 1.31 *Surface traps at the surface of a p-type semiconductor comprise electrons held in dangling bonds. The energy needed to release these electrons is approximately $\frac{E_g}{2}$. Since there are large numbers of dangling bond states, some being occupied and some not being occupied by electrons, the Fermi energy becomes pinned at this energy*

dangling bonds therefore comprise electrons that are likely to lie at approximately mid-gap because the energy required to excite them into the conduction band is only approximately half as large as the energy E_g required to remove an electron from the complete covalent bond of the relevant perfect crystal.

Now consider a large number of such dangling bonds at a semiconductor surface. Some of these dangling bonds will have lost electrons and some of them will not. Since the Fermi level exists between the highest filled states and the lowest empty state, the Fermi level tends to fall right onto the energy level range of these traps. The Fermi energy gets *pinned* to this trap energy at $\cong \frac{E_g}{2}$. Figure 1.31 shows the pinning of a Fermi level due to surface traps in a p-type semiconductor. Notice that at the semiconductor surface the surface traps determine the position of the Fermi energy rather than the doping level. An electric field is established in the semiconductor normal to the surface and *band bending* occurs as shown.

If the semiconductor had been n-type instead of p-type, then the same reasoning would still pin the Fermi energy to mid-gap; however, the band bending would occur in the opposite direction and the resulting electric field would point in the opposite direction, as illustrated in Figure 1.32.

Since traps are often formed from defects other than free surfaces that also involve incomplete bonding like a vacancy, a dislocation line, a grain boundary or an interface between two layers, this simple picture is very useful and will be used in the context of the p-n junction to explain recombination processes in subsequent sections of this book.

There is a velocity associated with excess minority carriers at a semiconductor surface or an interface between a semiconductor and another material. For example in Figure 1.31, which shows a p-type semiconductor, assume that there are excess electrons $\delta n(x)$ in the conduction band. The conduction band excess electron concentration decreases towards the
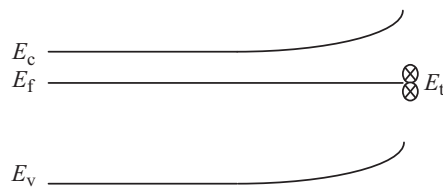

Figure 1.32 *Surface traps at the surface of an n-type semiconductor causing the Fermi level to be trapped at approximately mid-gap. An electric field opposite in direction to that of Figure 1.28 is formed in the semiconductor*

surface. This causes electrons to diffuse towards the surface, where they recombine. The diffusion of electrons towards the surface can be expressed as an electron flux

$$\varphi = -D_{\mathrm{n}}\frac{\mathrm{d}\delta n}{\mathrm{d}x}$$

Flux has units of particles per unit area per unit time. Since we can equivalently define a flux as the product of concentration and velocity, we can write

$$\varphi = -D_{\mathrm{n}}\frac{\mathrm{d}\delta n}{\mathrm{d}x} = S_{\mathrm{n}}\delta n \tag{1.66a}$$

where S_{n} is the *surface recombination velocity* of the electrons, and we evaluate δn and $\frac{\mathrm{d}\delta n}{\mathrm{d}x}$ at the semiconductor surface. Note that if $S_{\mathrm{n}} = 0$, we can conclude that $\frac{\mathrm{d}\delta n}{\mathrm{d}x}$ is zero, there is no band bending and surface states do not form a charged layer. Conversely if $S_n \to \infty$ then $\delta n = 0$ at the surface, which implies that carriers very rapidly recombine at the surface. A similar situation exists at the surface or at an interface of an n-type semiconductor in which holes may recombine and we obtain

$$\varphi = -D_{\mathrm{p}}\frac{\mathrm{d}\delta p}{\mathrm{d}x} = S_{\mathrm{p}}\delta p \tag{1.66b}$$

In both Equations 1.66a and 1.66b we assume that the excess carrier concentrations are much larger than the equilibrium minority carrier concentrations.

1.20   Alloy Semiconductors

An important variation in semiconductor compositions involves the use of partial substitutions of elements to modify composition. One example is the partial substitution of germanium in silicon that results in a range of new semiconductors of composition $Si_{1-x}Ge_x$, which are known as *alloy semiconductors*. The germanium atoms randomly occupy lattice sites normally occupied by silicon atoms, and the crystal structure of silicon is maintained. Note that Si and Ge are both in the group IV column of the periodic table and therefore have chemical similarities in terms of valence electrons and types of bonding. This means that provided no additional dopant impurities are introduced into the alloy semiconductor, alloy material with characteristics of an intrinsic semiconductor can be achieved. Of interest in semiconductor devices is the opportunity to modify the optical and electrical properties of the semiconductor. Since germanium has a smaller bandgap than silicon, adding germanium decreases bandgap as x increases. In addition, the average lattice constant of the new compound will increase since germanium is a larger atom than silicon. Since both germanium and silicon have the same diamond crystal structure, the available range of x is from 0 to 1 and the indirect bandgaps of the alloy compositions $Si_{1-x}Ge_x$ therefore range between 1.11 eV and 0.067 eV as x varies from 0 to 1 respectively.

Of more relevance to p-n junctions for solar cells and LEDs, alloy semiconductors may also be formed from compound semiconductors. For example $Ga_{1-x}In_xN$ is a *ternary*, or three-component, alloy semiconductor in which a fraction of the gallium atoms in wurtzite GaN is replaced by indium atoms. The indium atoms randomly occupy the crystalline sites in GaN that are normally occupied by gallium atoms. Since In and Ga are both group

III elements, the substitution does not act as either an acceptor or a donor. The direct bandgap decreases as x increases. For $x = 0$, $E_g = 3.4$ eV and for $x = 1$, $E_g = 0.77$ eV.

There are many other III-V alloy semiconductors. In $Ga_{1-x}Al_xAs$ alloys the bandgap varies from 1.43 to 2.16 eV as x goes from 0 to 1. In this system, however, the bandgap is direct in the case of GaAs, but indirect in the case of AlAs. There is a transition from direct to indirect bandgap at $x \cong 0.38$. We can understand this transition if we consider the two conduction band minima in GaAs shown in Figure 1.16c. One minimum forms a direct energy gap with the highest energy levels in the valence band; however, the second minimum forms an indirect gap. When $x \cong 0.38$ these two minima are at the same energy level. For $x < 0.38$ the alloy has a direct gap because the global conduction band minimum forms the direct gap. For $x > 0.38$ the global conduction band minimum is the minimum that forms the indirect gap.

A number of III-V alloy systems are illustrated in Figure 1.33a. An additional set of III-V nitride semiconductors is shown in Figure 1.33b, and a set of II-VI semiconductors is included in Figure 1.33c.
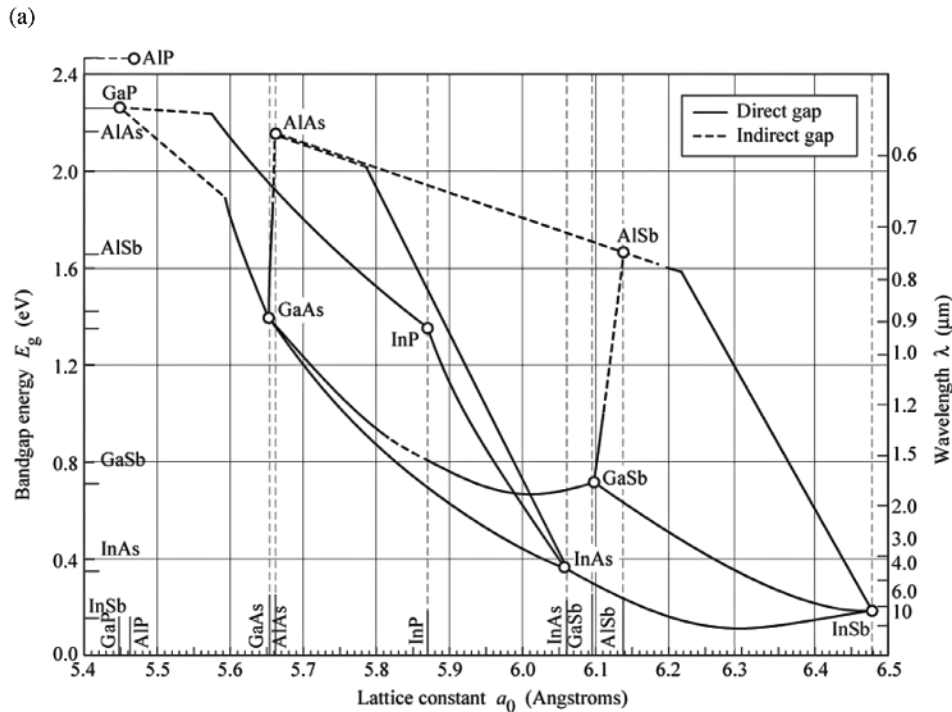

Figure 1.33 *Bandgap versus lattice constant for (a) phosphide, arsenide and antimonide III-V semiconductors; (b) nitride and other III-V semiconductors neglecting bowing (see Figure 5.20); (c) sulphide, selenide and telluride II-VI semiconductors and phosphide, arsenide and antimonide III-V semiconductors. (a–b) Reprinted from E. Fred Schubert, Light-Emitting Diodes, 2e ISBN 978-0-521-86538-8. Copyright (2006) with permission from E. Fred Schubert. (c) Adapted from http://www.tf.uni-kiel.de/matwis/amat/semitech_en/ Copyright Prof. Dr. Helmut Föll, Technische Fakultät, Universität Kiel*
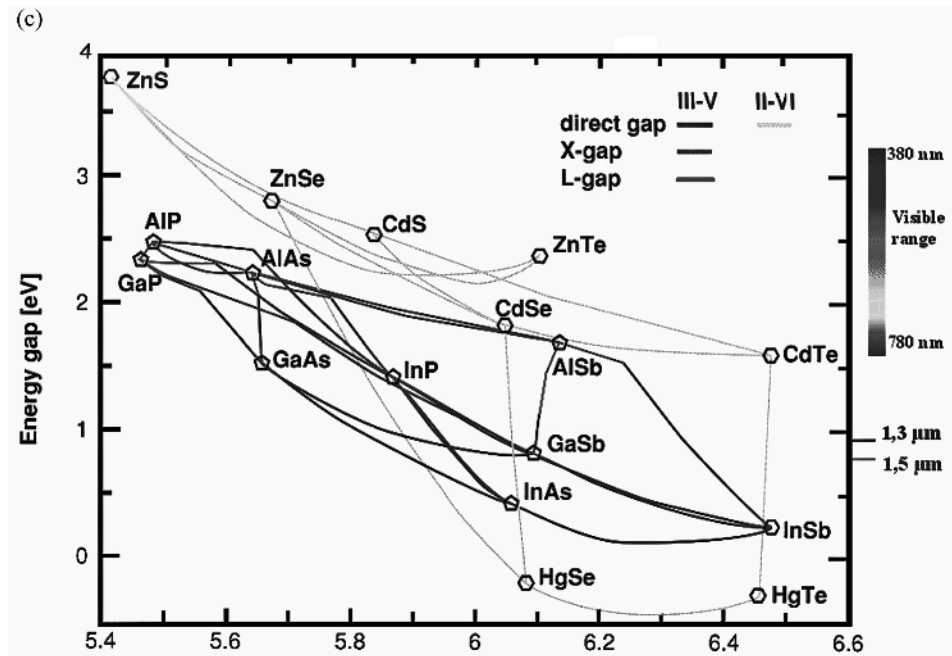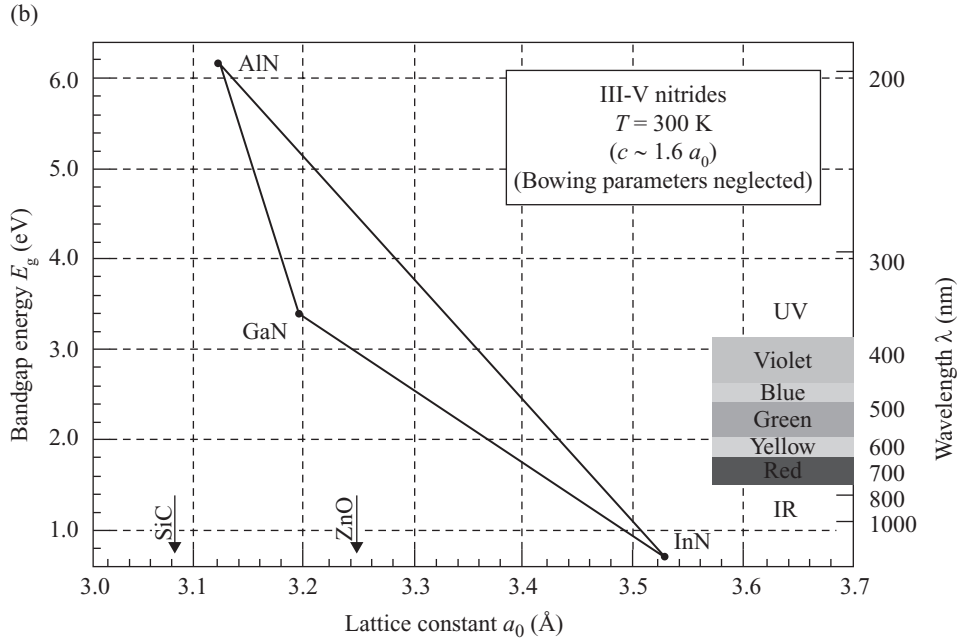
(b)

(c)

Figure 1.33   *(Continued)*

To cover ranges of composition in the III-V alloy system, *quaternary* alloys may be formed such as $(Al_xGa_{1-x})_yIn_{1-y}P$. The use of ternary and quaternary semiconductor alloys in solar cells and LEDs will be discussed in Chapters 4 and 5.

1.21   Summary

1.1. The band theory of solids permits an understanding of electrical and optical properties including electrical conductivity in metals and semiconductors, optical absorption and luminescence, and properties of junctions and surfaces of semiconductors and metals.

1.2. The band theory of solids requires the use of wavefunctions to describe the spatial distribution, energy and momentum of electrons. Schrödinger's equation may be solved to obtain these wavefunctions by employing a suitable potential energy function in this equation.

1.3. In the Kronig–Penney model a periodic potential leads to energy bands, energy gaps and Brillouin zones by solving Schrödinger's equation for electrons in a periodic potential. The size of the energy gaps increases as the amplitude of the periodic potential increases. As the ionic character of the bonding in the semiconductor increases the energy gap increases. As the size of the atoms decreases the energy gap increases.

1.4. The Bragg model identifies Brillouin zone boundaries as satisfying the Bragg condition for strong reflection. This condition is $2a = n\lambda$. The reduced zone scheme, which shows only the first Brillouin zone, simplifies the representation of energy bands and energy gaps.

1.5. The effective mass m^* is used to quantify electron behaviour in response to an applied force. The effective mass depends on the band curvature. Effective mass is constant if the band shape can be approximated as parabolic.

1.6. The number of states in a band n can be determined based on the number of unit cells N in the semiconductor sample. The result $n = 2N$ is obtained for a one-dimensional, two-dimensional or three-dimensional case.

1.7. The filling of bands in semiconductors and insulators is such that the highest filled band is full and the lowest empty band is empty at low temperatures. In metals the highest filled band is only partly filled. Semiconductors have smaller bandgaps ($E_g = 0$ to $4\,eV$) than insulators ($E_g > 4\,eV$).

1.8. The Fermi energy E_f is defined as the energy level at which an electron state has a 50% probability of occupancy at temperatures above $0\,K$. A hole can be created when an electron from the valence band is excited to the conduction band. The hole can move independently from the electron.

1.9. Carrier concentration in an energy band is determined by (i) finding the probability of occupancy of the states in a band using the Fermi–Dirac distribution function $F(E)$, and (ii) finding the density of states function $D(E)$ for an energy band. Then the integral over the energy range of the band of the product of $D(E)$ and $F(E)$ will determine the number of carriers in the band. In the conduction band the equilibrium electron concentration is n_0 and in the valence band the equilibrium hole concentration is p_0. The product $n_0 p_0$ is a constant that is independent of the Fermi energy.

1.10. A range of semiconductor materials includes group IV semiconductors, group III-V semiconductors and group II-VI semiconductors as listed in order of increasing ionic character. Bandgap energies decrease for larger atoms that are lower down on the periodic table.

1.11. Most important semiconductor crystals have lattices that are diamond, zincblende or hexagonal structures. These structures lead to complex band diagrams. The band shapes in E versus k plots depend on crystallographic directions. In addition there are sub-bands in the valence band that correspond to distinct hole effective masses. Bandgaps may be direct or indirect. Conduction bands generally exhibit two minima where one minimum corresponds to a direct gap transition and one minimum corresponds to an indirect gap transition.

1.12. Photon momentum is very small and direct gap transitions in semiconductors are favourable for photon creation and absorption. In indirect gap transitions the involvement of lattice vibrations or phonons is required. The absorption coefficient α is higher for direct gap semiconductors and lower for indirect gap semiconductors for photons above E_g in energy. Whereas indirect gap silicon has an effective absorption depth of $\cong 100\,\mu m$ for sunlight the corresponding absorption depth in GaAs is only $\cong 1\,\mu m$.

1.13. Pure semiconductors are known as intrinsic semiconductors. The incorporation of low levels of impurity atoms in a semiconductor leads to extrinsic semiconductors, in which the electron concentration n_0 and hole concentration p_0 are controlled by the impurity type and concentration. Donor impurities donate electrons to the conduction band in n-type semiconductors, and acceptor impurities donate holes to the valence band in p-type semiconductors. New shallow energy levels arise within the energy gap, which are called donor and acceptor levels. Carrier concentrations are temperature dependent; however, over a wide intermediate temperature range carrier concentrations are relatively constant as a function of temperature. Minority carriers refer to the carriers having a low concentration in a specific semiconductor region, and majority carriers refer to the carriers having a significantly higher concentration in the same region.

1.14. Carriers move through semiconductors in an electric field ε by a drift process, which is characterized by a drift velocity $\bar{v}$ and mobility μ. Drift current density is given by $J = \sigma\varepsilon$, which is an expression of Ohm's law. The understanding of Ohm's law is based on the concept of a terminal velocity due to scattering events having a characteristic scattering time τ, which depends on impurities, defects and temperature. At high electric fields drift velocity will eventually saturate.

1.15. Carrier concentrations are not necessarily at equilibrium levels. Photons or applied electric fields can give rise to non-equilibrium excess carrier concentrations, which will return to equilibrium concentrations once equilibrium conditions are restored. Electron-hole pair (EHP) generation and recombination processes G_{th} or G_{op} and R define the resulting rate of generation and recombination. Under equilibrium conditions $G_{th} = R$ where $R \propto np$. The minority carrier lifetime time constants τ_n or τ_p characterize the recombination times of minority carriers.

1.16. Carriers diffuse in semiconductors due to a concentration gradient. The diffusion coefficients D_n and D_p determine the diffusion current J_n and J_p respectively. The net current flow must include both drift and diffusion current. In equilibrium the net

current is zero; however, drift and diffusion currents may be non-zero. The Einstein relation is derived from the requirements for equilibrium conditions and allows D_n and D_p to be derived from μ_n and μ_p.

1.17. In non-equilibrium conditions the Fermi energy is not defined; however, quasi-Fermi energies F_n and F_p may be defined to characterize changes in carrier concentrations due to excess carrier generation.

1.18. Combining the concepts of carrier recombination and carrier diffusion the diffusion equation leads to the calculation of diffusion length $L_n = \sqrt{D_n \tau_n}$ and $L_p = \sqrt{D_p \tau_p}$ and shows that carrier concentration decays exponentially as a function of distance from a region of excess carrier generation.

1.19. Traps can have a large effect on carrier lifetimes. The most important traps are deep traps that are at or near mid-gap. A high density of such traps occurs at semiconductor interfaces and defects. This is due to dangling bonds at surfaces and defects. Fermi level pinning occurs at or near midgap due to dangling bonds. Surface recombination velocity is a measure of the rate of recombination at semiconductor surfaces.

1.20. The bandgap and direct/indirect nature of a semiconductor can be altered by alloying. Alloying may also change the lattice constant. Industrially important alloy semiconductors exist composed from group IV elements, group III-V elements and group II-VI elements. Ternary and quaternary compound semiconductors are important for solar cells and LEDs.

Suggestions for Further Reading

Ashcroft NW and Mermin ND. *Solid State Physics*. Holt, Rinehart and Winston, 1976.

Kittel C. *Introduction to Solid State Physics*, 8th edn. John Wiley and Sons, 2005.

Eisberg R and Resnick R. *Quantum Physics of Atoms, Molecules, Solids, Nuclei and Particles*, 2nd edn. John Wiley and Sons, 1985.

Neamen DA. *Semiconductor Physics and Devices*, 3rd edn. McGraw Hill, 2003.

Solymar L and Walsh D. *Electrical Properties of Materials*, 7th edn. Oxford University Press, 2004.

Problems

1.1 Derive the following relationships in the Kronig–Penney model:
  (a) Obtain Equation 1.7c from Equations 1.4a, 1.4b, 1.7a and 1.7b.
  (b) Obtain Equation 1.8 from Equation 1.7c. Take the limit $b \to 0$ and $V_0 \to \infty$ such that bV_0 is constant.
  Use the definition

$$P = \frac{Q^2 ba}{2}$$

1.2 Using the Kronig–Penney model in Equation 1.8:
  (a) If $P = 2$ and $a = 2.5$ Å find the energy range for the two lowest energy bands and the energy gap between them.
  (b) Repeat (a) if $P = 0.2$.
  (c) Demonstrate that the energy bands reduce to free electron states as $P \to 0$.

1.3 In Section 1.6 we showed that the number of states in an energy band is $n = 2N$ for a one-dimensional semiconductor. Show that the number of states in an energy band in a three-dimensional semiconductor is still $n = 2N$ where N is the number of unit cells in the three-dimensional semiconductor.
  Hint: Consider a semiconductor in the form of a rectangular box having N_x, N_y and N_z unit cells along the x, y and z axes. Assume an infinite walled box with $V = 0$ inside the box. Use the allowed energy values for an electron in Equation 1.28. Include spin.

1.4 A rectangular semiconductor crystal has dimensions $2 \times 2 \times 1$ mm. The unit cell is cubic and has edge length of 2 Å. Find the number of states in one band of this semiconductor.

1.5 A rectangular silicon semiconductor bar of length 12 cm and cross-section 1×5 mm is uniformly doped n-type with concentration $N_d = 5 \times 10^{16}$ cm^{-3}.
  (a) Assuming all donors are ionized, calculate the room temperature current flow if contacts are made on the two ends of the bar and 10 V is applied to the bar.
  (b) Find the electric field in the bar for the conditions of (a).
  (c) What fraction of the current flows in the form of hole current for the conditions of (a)?
  (d) Find the resistivity of the silicon.
  (e) If the silicon were replaced by gallium phosphide and the doping was still $N_d = 5 \times 10^{16}$ cm^{-3} repeat (a), (b) and (c).
  (f) If the silicon temperature was increased to 120°C, repeat (a), (b) and c). Assume that carrier mobility and bandgap are not affected by the increase in temperature.

1.6 Now, instead of being uniformly doped, the bar of Problem 1.5 is doped with a linearly increasing donor doping concentration, such that the left end of the bar (LHS) is doped with a concentration of 1×10^{16} cm^{-3} and the right end of the bar (RHS) is doped with a concentration of 1×10^{17} cm^{-3}.
  (a) Determine the doping level at three points in the bar:
    (i) at 3 cm from the LHS;
    (ii) at the midpoint;
    (iii) at 9 cm from the LHS.

(b) Assuming equilibrium conditions (no applied voltage) find the built-in electric field in the bar at each of positions (i), (ii) and (iii). Hint: Find the gradient in the doping about each point.

(c) Find the electron drift current flowing in the bar at positions (i), (ii) and (iii).

(d) Explain how the bar can be in equilibrium given the existence of these electric fields and drift currents.

(e) Find the Fermi level relative to the top of the valence band for each of positions (i), (ii) and (iii).

(f) Sketch the band diagram as a function of position in equilibrium along the length of the bar showing the location of the Fermi energy.

1.7 A square silicon semiconductor sheet $50 \, \text{cm}^2$ in area and $0.18 \, \text{mm}$ in thickness is uniformly doped with both acceptors ($N_a = 5 \times 10^{16} \, \text{cm}^{-3}$) and donors ($N_d = 2 \times 10^{16} \, \text{cm}^{-3}$).

(a) Assuming all donors and acceptors are ionized, calculate the room temperature current flow if the silicon is contacted by metal contact strips that run the full length of two opposing edges of the sheet and $10 \, \text{V}$ is applied across the contacts.

(b) Repeat (a) but assume that the sheet is $100 \, \text{cm}^2$ in area instead of $50 \, \text{cm}^2$. Does the current change with area? Explain.

(c) Is the recombination time τ important for the measurement of current flow? Explain carefully what role recombination plays in the calculation.

1.8 An undoped silicon semiconductor sheet $50 \, \text{cm}^2$ in area and $0.18 \, \text{mm}$ thick is illuminated over one entire surface and an electron-hole pair generation rate of $10^{21} \, \text{cm}^{-3} \, \text{s}^{-1}$ is achieved uniformly throughout the material.

(a) Determine the separation of the quasi-Fermi levels. The carrier lifetime is $2 \times 10^{-6} \, \text{s}$.

(b) Calculate the room temperature current flow if the silicon is contacted by narrow metal contact strips that run the full length of two opposing edges of the sheet and $10 \, \text{V}$ is applied across the contacts under illumination conditions. The carrier lifetime is $2 \times 10^{-6} \, \text{s}$.

(c) Explain how a higher/lower recombination time would affect the answer to (b). How does the recombination time of $2 \times 10^{-6} \, \text{s}$ compare with the transit time of the carriers, which is the time taken by the carriers to traverse the silicon sheet from one side to the other side? This silicon sheet is functioning as a *photoconductive* device since its conductivity depends on illumination. If the transit time is small compared to the recombination time then *gain* can be obtained since more than one carrier can cross the photoconductive sheet before a recombination event takes place on average. Gains of 100 or 1000 may be obtained in practice in photoconductors. How long a recombination time would be required for a gain of 10 to be achieved?

1.9 A sample of n-type silicon is doped to achieve E_F at $0.3 \, \text{eV}$ below the conduction band edge at room temperature.

(a) Find the doping level.

(b) The n-type silicon sample is in the form of a square cross-section bar at room temperature and it carries a current of $3 \times 10^{-8} \, \text{Å}$ along its length. If the bar is

10 cm long and has a voltage difference of 100 V end-to-end, find the cross-section dimension of the bar.

1.10 A silicon sample is uniformly optically excited such that its quasi-Fermi level for electrons F_n is 0.419 eV above its quasi-Fermi level for holes F_p. The silicon is n-type with donor concentration $N_D = 1 \times 10^{14}$ cm^{-3}. Find the optical generation rate.

1.11 An n-type silicon wafer is 5.0 mm thick and is illuminated uniformly over its surface with blue light, which is absorbed very close to the silicon surface. Assume that a *surface generation rate* of holes of 3×10^{18} cm^{-2} s^{-1} is obtained over the illuminated surface, and that the excess holes are generated at the silicon surface.

  (a) Calculate the hole concentration as a function of depth assuming a hole lifetime of 2×10^{-6} s. Assume that the hole lifetime is independent of depth.
  (b) Calculate the hole diffusion current as a function of depth.
  (c) Calculate the recombination rate of holes as a function of depth.
  (d) Explain the difference between the surface generation rate of holes and the recombination rate of holes near the surface. These two rates have different units. Your explanation should refer to the distinction between the units employed.

1.12 Find:

  (a) The n-type doping level required to cause silicon at room temperature to have electrical conductivity 100 times higher than intrinsic silicon at room temperature.
  (b) The p-type doping level required to cause silicon at room temperature to have p-type conductivity 100 times higher than intrinsic silicon at room temperature.

1.13 Intrinsic silicon is uniformly illuminated with 10^{14} photons cm^{-2} s^{-1} at its surface. Assume that each photon is absorbed very near the silicon surface, and generates one electron-hole pair.

  (a) Find the flux of electrons at a depth of 3 μm. Make and state any necessary assumptions.
  (b) Find the total excess electron charge stored in the silicon, assuming the silicon sample is very thick.

1.14 If the Fermi energy in an n-type silicon semiconductor at 300°C is 0.08 eV below the conduction band, and the donor level is 0.02 eV below the conduction band, then find the probability of ionization of the donors.

1.15 In a hypothetical semiconductor with effective masses of electrons and holes the same as silicon, and E_g as in silicon, the conductivity of a rod of the material (measuring 1 mm in diameter and 40 mm in length) from end to end is measured as 4×10^{12} ohms at 300 K.

  (a) If the electron mobility is five times larger than the hole mobility in this material then find the electron mobility. Make and state any necessary assumptions.
  (b) Find the current flow due to the holes only in the rod if a voltage of 10 V is applied across the ends of the rod. Make and state any necessary assumptions.

1.16 An intrinsic, planar, room-temperature silicon sample is exposed to a steady flux of light at its surface. The electron concentration as a result of this is measured to be 100 times higher than n_i, the intrinsic equilibrium concentration, at a depth of 100 μm below the silicon surface. You may use the low-level injection approximation.

  (a) Assuming that the light is all absorbed very near the silicon surface, and that every incident photon excites one EHP, find the total photon flux.

(b) Find the quasi-Fermi level for electrons relative to the Fermi level without the flux of light at a depth of 100 μm.

1.17 A famous experiment that involves both the drift and the diffusion of carriers in a semiconductor is known as the Haynes–Shockley experiment. Search for the experimental details of this experiment and answer the following:

(a) Make a sketch of the semiconductor sample used in the experiment as well as the location and arrangement of electrodes and the required voltages and currents as well as the connections of electrodes for the appropriate measurements to be made.

(b) Sketch an example of the time dependence of the output of the experiment.

(c) Explain how the Einstein relation can be verified using these data.

1.18 A silicon sample is uniformly optically excited such that its quasi-Fermi level for electrons F_n is 0.419 eV above its quasi-Fermi level for holes F_p. The silicon is n-type with donor concentration $N_d = 1 \times 10^{17}$ cm^{-3}. Find the optical generation rate.

1.19 A flash of light at time $t = 0$ is uniformly incident on all parts of a p-type silicon sample with doping of 5×10^{17} cm^{-3}. The resulting EHP concentration is 2×10^{16} EHP cm^{-3}. Find the time-dependent electron and hole concentrations for time t greater that zero.

1.20 Carriers are optically generated at an intrinsic silicon surface. The generation rate is 2×10^{19} EHP/cm^2 s. Assume that all the photons are absorbed very close to the silicon surface. Find the diffusion current of electrons just below the surface, and state clearly the assumptions you used to obtain the result.

1.21 A silicon sample is doped with 6×10^{16} donors cm^{-3} and N_a acceptors cm^{-3}. If E_f lies 0.4 eV below E_f in intrinsic silicon at 300 K, find the value of N_a.

1.22 Electric current flows down a silicon rod 1 cm in length and 0.3 mm in diameter. The silicon is n-type with $N_a = 1 \times 10^{17}$ cm^{-3}. A potential difference of 10 V is applied to the rod end-to-end. How many electrons drift through the rod in 60 seconds?

1.23 Find the energy difference between F_n (quasi-Fermi level for electrons) and E_f if illumination were not present, for a silicon sample in equilibrium containing 10^{15} donors cm^{-3}. Assume room temperature. The bar is uniformly optically excited such that 10^{18} EHP/cm^3 s are generated.

1.24 (a) Find the surface recombination velocity of electrons at a p-type silicon semiconductor surface having the following parameters:

$$p_0 = 1 \times 10^{17} \text{ cm}^{-3}$$
$$\delta n = 1 \times 10^{16} \text{ cm}^{-3} \text{ at the surface.}$$

Use the diffusion constant for silicon in Appendix 2. Make and state any necessary assumptions.

(b) Sketch a band diagram as a function of distance x from the semiconductor surface to a few diffusion lengths away from the surface.

(c) Repeat (a) and (b) for a GaAs semiconductor with the same parameters.

1.25 (a) Find the surface recombination velocity of holes at an n-type silicon semiconductor surface having the following parameters:

$$n_0 = 1 \times 10^{16} \text{ cm}^{-3}$$

$$\delta p = 1 \times 10^{17} \text{ cm}^{-3} \text{ at the surface.}$$

Use the diffusion constant for silicon in Appendix 2. Make and state any necessary assumptions.

(b) Sketch a band diagram as a function of distance x from the semiconductor surface to a few diffusion lengths away from the surface.

(c) Repeat (a) and (b) for a GaAs semiconductor with the same parameters.