1

Electrons and Holes in Semiconductors

CHAPTER OBJECTIVES

This chapter provides the basic concepts and terminology for understanding semiconductors. Of particular importance are the concepts of energy band, the two kinds of electrical charge carriers called electrons and holes, and how the carrier concentrations can be controlled with the addition of dopants. Another group of valuable facts and tools is the Fermi distribution function and the concept of the Fermi level. The electron and hole concentrations are closely linked to the Fermi level. The materials introduced in this chapter will be used repeatedly as each new device topic is introduced in the subsequent chapters. When studying this chapter, please pay attention to (1) concepts, (2) terminology, (3) typical values for Si, and (4) all boxed equations such as Eq. (1.7.1).

The title and many of the ideas of this chapter come from a pioneering book, *Electrons and Holes in Semiconductors* by William Shockley [1], published in 1950, two years after the invention of the transistor. In 1956, Shockley shared the Nobel Prize in physics for the invention of the transistor with Brattain and Bardeen (Fig. 1–1).

The materials to be presented in this and the next chapter have been found over the years to be useful and necessary for gaining a deep understanding of a large variety of semiconductor devices. Mastery of the terms, concepts, and models presented here will prepare you for understanding not only the many semiconductor devices that are in existence today but also many more that will be invented in the future. It will also enable you to communicate knowledgeably with others working in the field of semiconductor devices.

1.1 ● SILICON CRYSTAL STRUCTURE ●

A crystalline solid consists of atoms arranged in a repetitive structure. The periodic structure can be determined by means of X-ray diffraction and electron microscopy. The large cubic unit shown in Fig. 1–2 is the **unit cell** of the silicon

● **Inventors of the Transistor** ●

Born on three different continents (Brattain in Amoy, China; Bardeen in Madison, Wisconsin, USA; and Shockley in London, England), they all grew up in the United States and invented the transistor in 1947–1948 at Bell Telephone Laboratories. Brattain was an experimentalist while Bardeen and Shockley contributed more to the concepts and theories. Their reflections on that historic event:

"... after fourteen years of work, I was beginning to give up ..."

—Walter H. Brattain (1902–1987)

"Experiments that led to the invention of the point-contact transistor by Walter Brattain and me were done in November and December, 1947, followed closely by the invention of the junction transistor by Shockley."

—John Bardeen (1908–1991)

"All of us who were involved had no doubt that we had opened a door to a new important technology. "
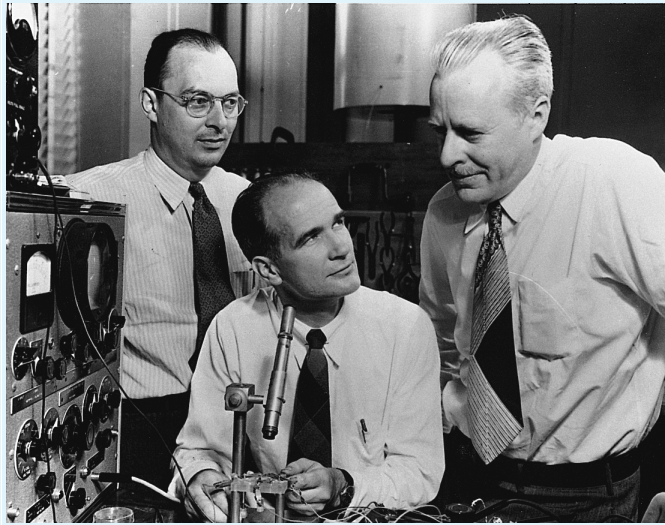
—William B. Shockley (1910–1988)

FIGURE 1–1 Transistor inventors John Bardeen, William Shockley, and Walter Brattain (left to right) at Bell Telephone Laboratories. (Courtesy of Corbis/Bettmann.)

crystal. Each sphere represents a silicon atom. This unit cell is repeated in all three directions many times to form a silicon crystal. The length of the unit cell, e.g., 5.43 Å in Fig. 1–2, is called the **lattice constant**.

The most important information from Fig. 1–2 is the simple fact that *each and every silicon atom has four other silicon atoms as its nearest neighbor atoms*. This fact is illustrated in Fig. 1–2 with the darkened cluster of a center atom having four neighboring atoms. This cluster is called the **primitive cell**. Silicon is a group IV element in the periodic table and has four valence electrons. These four electrons are shared with the nearest neighbors so that eight covalent electrons are associated
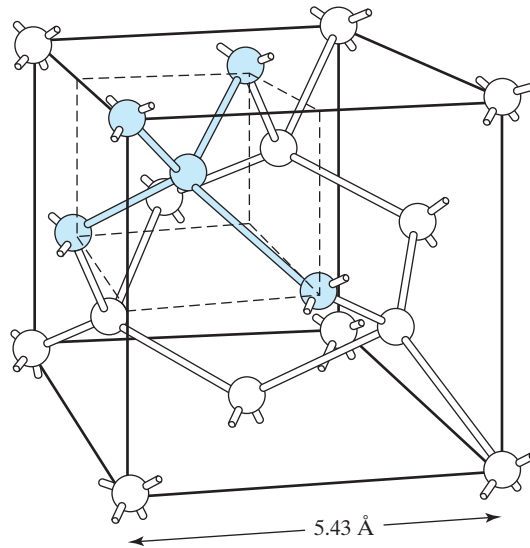
FIGURE 1–2 The unit cell of the silicon crystal. Each sphere is a Si atom. Each Si atom has four nearest neighbors as illustrated in the small cube with darkened atoms. (Adapted from Shockley [1].) For an interactive model of the unit cell, see http://jas.eng.buffalo.edu/

with each atom. The structure shown in Fig. 1–2 is known as the **diamond structure** because it is also the unit cell of the diamond crystal with each sphere representing a carbon atom. Germanium, the semiconductor with which the first transistor was made, also has the diamond crystal structure.

Figure 1–3 introduces a useful system of denoting the orientation of the silicon crystal. The cube in Fig. 1–3a represents the Si unit cell shown in Fig. 1–2 and each darkened surface is a crystal plane. The (100) crystal plane in the leftmost drawing in Fig. 1–3a, for example, is simply the plane in Fig. 1–2 closest to the reader. It intersects the x axis at 1 lattice constant and the y and z axes at infinity. One might refer to this plane as the 1 ∞ ∞ plane. However, it is standard practice to refer to it as the (1/1 1/∞ 1/∞), or the (100), plane. In general, the (abc) plane intersects the x, y, and z axes at $1/a$, $1/b$, and $1/c$ lattice constants. For example, the (011) plane in the middle drawing in Fig. 1–3a intersects the x axis at infinity and the y and z axes at 1 lattice constant. The numerals in the parentheses are called the **Miller indices**. The related symbol $[abc]$ indicates the direction in the crystal normal to the (abc) plane. For example, when an electron travels in the [100] direction, it travels perpendicular to the (100) plane, i.e., along the x axis.

Figure 1–3b shows that the silicon wafers are usually cut along the (100) plane to obtain uniformity and good device performance. A flat or a notch is cut along the (011) plane in order to precisely and consistently orient the wafer as desired during device fabrication. Different surface orientations have different properties such as the rate of oxidation and the electronic quality of the oxide/semiconductor interface. Both the surface orientation and the direction of current flow along the surface affect the speed performance of a surface-base device such as metal-oxide-semiconductor field-effect transistor (MOSFET, see Section 6.3.1). The most important semiconductor materials used in microelectronics are crystalline. However, most everyday solids are not single crystals as explained in the sidebar in Section 3.7.
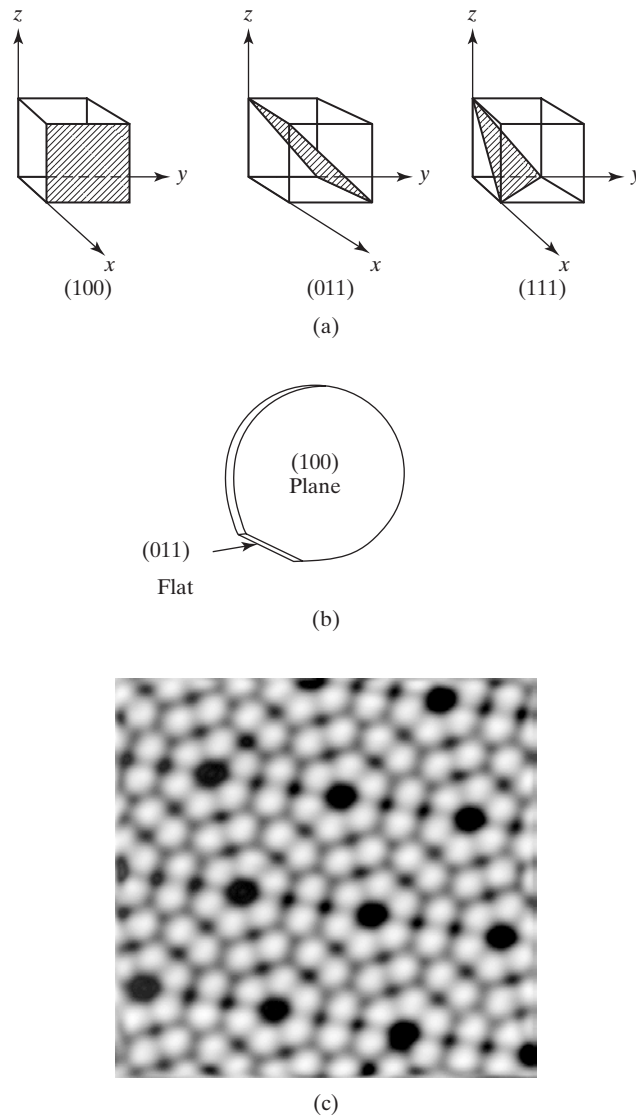
FIGURE 1–3 (a) A system for describing the crystal planes. Each cube represents the unit cell in Fig. 1–2. (b) Silicon wafers are usually cut along the (100) plane. This sample has a (011) flat to identify wafer orientation during device fabrication. (c) Scanning tunneling microscope view of the individual atoms of silicon (111) plane.

1.2   ●   BOND MODEL OF ELECTRONS AND HOLES   ●

Each silicon atom is surrounded by four nearest neighbors as illustrated by the shaded cluster in Fig. 1–2. We can represent the silicon crystal structure with the two-dimensional drawing shown in Fig. 1–4. An Si atom is connected to each neighbor with two dots representing the two shared electrons in the covalent bond. Figure 1–4 suggests that there are no free electrons to conduct electric current. This is strictly true

$$: \overset{\cdot\,\cdot}{\underset{\cdot\,\cdot}{Si}} : \overset{\cdot\,\cdot}{\underset{\cdot\,\cdot}{Si}} : \overset{\cdot\,\cdot}{\underset{\cdot\,\cdot}{Si}} :$$
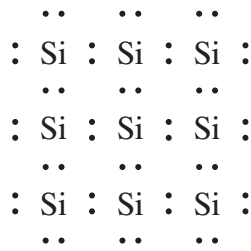
FIGURE 1–4 The silicon crystal structure in a two-dimensional representation.

only at the absolute zero temperature. At any other temperature, thermal energy will cause a small fraction of the covalent electrons to break loose and become **conduction electrons** as illustrated in Fig. 1–5a. Conduction electrons can move around in a crystal and therefore can carry electrical currents. For this reason, the conduction electrons are of more interest to the operation of devices than valence electrons.

An interesting thing happens when an electron breaks loose and becomes free. It leaves behind a void, or a **hole** indicated by the open circle in Fig. 1–5a. The hole can readily accept a new electron as shown in Fig. 1–5b. This provides another means for electrons to move about and conduct currents. An alternative way to think of this process is that the hole moves to a new location. It is much easier to think of this second means of current conduction as the motion of a positive hole than the motion of negative electrons moving in the opposite direction just as it is much easier to think about the motion of a bubble in liquid than the liquid movement that creates the moving bubble.

In semiconductors, current conduction by holes is as important as electron conduction in general. *It is important to become familiar with thinking of the holes as mobile particles carrying positive charge, just as real as conduction electrons are mobile particles carrying negative charge.* It takes about 1.1 eV of energy to free a covalent electron to create a conduction electron and a hole. This energy can be determined, for example, from a photoconductivity experiment. When light shines on a Si sample, its conductivity increases because of the generation of mobile electrons and holes. The minimum photon energy required to induce photoconductivity is 1.1 eV.

The densities of thermally generated electrons and holes in semiconductors are generally very small at room temperature given that the thermal energy, kT, is 26 meV at room temperature. A much larger number of conduction electrons can be introduced if desired by introducing suitable impurity atoms—a process called **doping**
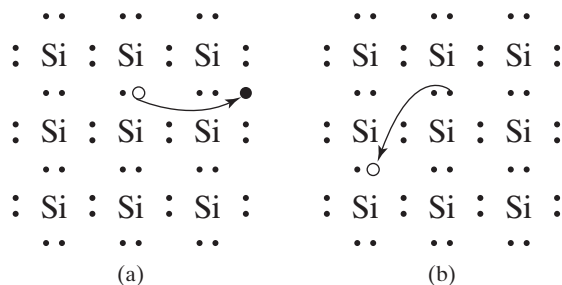
FIGURE 1–5 (a) When a covalent electron breaks loose, it becomes mobile and can conduct electrical current. It also creates a void or a hole represented by the open circle. The hole can also move about as indicated by the arrow in (b) and thus conduct electrical current.
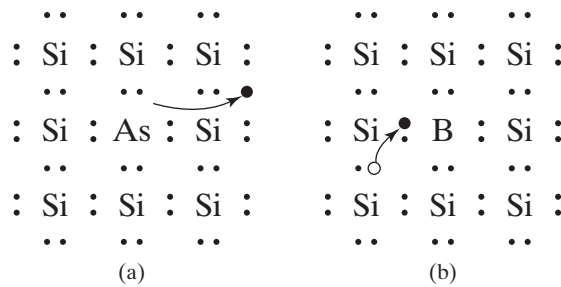
$$
\begin{array}{cccc}
\cdot\cdot & \cdot\cdot & \cdot\cdot & \\
: \text{Si} : \text{Si} : \text{Si} : & : \text{Si} : \text{Si} : \text{Si} : \\
\end{array}
$$

FIGURE 1–6 Doping of a semiconductor is illustrated with the bond model. (a) As is a donor. (b) B is an acceptor.

the semiconductor. For example, group V elements such as As shown in Fig. 1–6a bring five valence electrons with each atom. While four electrons are shared with the neighboring Si atoms, the fifth electron may escape to become a mobile electron, leaving behind a positive As ion. Such impurities are called **donors** for they *donate* electrons. Notice that in this case, no hole is created in conjunction with the creation of a conduction electron. Semiconductors containing many mobile electrons and few holes are called **N-type semiconductors** because electrons carry negative (N) charge. As and P are the most commonly used donors in Si.

Similarly, when boron, a group III impurity, is introduced into Si as shown in Fig. 1–6b, each B atom can accept an extra electron to satisfy the covalent bonds, thus creating a hole. Such dopants are called **acceptors**, for they *accept* electrons. Semiconductors doped with acceptors have many holes and few mobile electrons, and are called **P type** because holes carry positive (P) charge. Boron is the most commonly used acceptor in Si. In and Al are occasionally used.

The energy required to ionize a donor atom (i.e., to free the extra electron and leave a positive ion behind) may be estimated by modifying the theory of the ionization energy of a hydrogen atom,

$$
E_{\text{ion}} = \frac{m_0 q^4}{8\varepsilon_0^2 h^2} = 13.6 \text{ eV} \tag{1.2.1}
$$

where m_0 is the free electron mass, ε_0 is the permittivity of free space, and h is Planck's constant. The modification involves replacing ε_0 with $12\varepsilon_0$ (where 12 is the relative permittivity of silicon) and replacing m_0 with an electron effective mass, m_n, which is a few times smaller than m_0 as explained later. The result is about 50 meV. Because donors have such small ionization energies, they are usually fully ionized at room temperature. For example, 10^{17}cm^{-3} of donor atoms would lead to 10^{17}cm^{-3} of conduction electrons. The same conclusion also applies to the acceptors.

● GaAs, III–V Compound Semiconductors and Their Dopants ●

GaAs and similar **compound semiconductors**, such as InP and GaN, are dominant in optoelectronic devices such as light-emitting diodes and semiconductor lasers (see Sections 4.13 and 4.14). GaAs also plays a role in high-frequency electronics (see Sections 6.3.2 and 6.3.3). Its crystal structure is shown in Fig. 1–7 and Fig. 1–8. The similarity to the Si crystal is obvious. The shaded spheres represent As atoms and the

light spheres represent Ga atoms. Each Ga atom has four As neighbors and each As atom has four Ga neighbors. The lattice constant is 5.65 Å. Ga is a group III element and As is a group V element. GaAs is known as a **III–V compound semiconductor**, as are GaP and A1As, which also have the same crystal structure as illustrated in Fig. 1–7.

It is probably obvious that group VI elements such as S and Se can replace the group V As and serve as donors in GaAs. Similarly, group II elements such as Zn can replace Ga and serve as acceptors.

But, are group IV elements such as Si and Ge donors or acceptors in GaAs? The answer is that they can be either donors or acceptors, depending on whether they substitute for Ga atoms (which have three valence electrons) or As atoms (which have five valence electrons). Such impurities are called **amphoteric dopants**. It turns out that Si is a donor and Ge is an acceptor in GaAs because it is energetically more favorable for the small Si atoms to substitute for the small Ga atoms and for the larger Ge to substitute for the larger As.
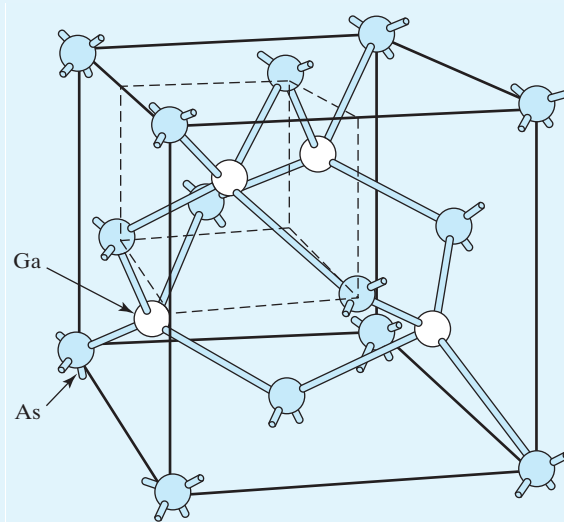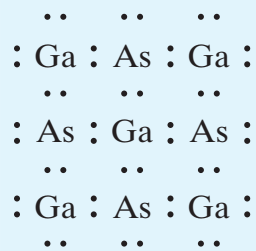

FIGURE 1–7 The GaAs crystal structure.

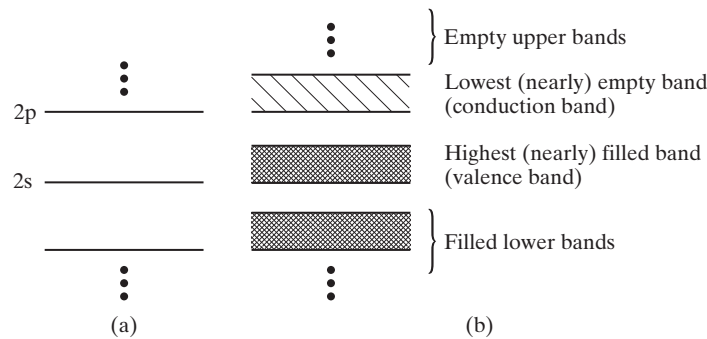

FIGURE 1–8 Bond model of GaAs.

FIGURE 1–9 The discrete energy states of a Si atom (a) are replaced by the energy bands in a Si crystal (b).

1.3 ● ENERGY BAND MODEL ●

While the bond model described in the previous section is conceptually simple, it is not complete enough for understanding semiconductor devices. The most useful model involves the concept of **energy bands**.

Recall that electrons in an atom occupy discrete energy levels as shown in Fig. 1–9a. If two atoms are in close proximity, each energy level will split into two due to the Pauli exclusion principle that states that each quantum state can be occupied by no more than one electron in an electron system such as an atom molecule, or crystal. When many atoms are brought into close proximity as in a crystal, the discrete energy levels are replaced with *bands* of energy states separated by gaps between the bands as shown in Fig. 1–9b. One may think of an energy band as a semicontinuum of a very large number of energy states.

Naturally, the electrons tend to fill up the low energy bands first. The lower the energy, the more completely a band is filled. In a semiconductor, most of the energy bands will be basically totally filled (completely filled at absolute zero), while the higher energy bands are basically totally empty. Between the (basically) totally filled and totally empty bands lie two bands that are only nearly filled and nearly empty as shown in Fig. 1–9b. They are of utmost interest to us. The top nearly filled band is called the **valence band** and the lowest nearly empty band is called the **conduction band**. The gap between them is called the **band gap**. The electrons in a totally filled band do not have a net velocity and do not conduct current, just as the water in a totally filled bottle does not slosh about. Similarly, a totally empty band cannot contribute to current conduction. These are the reasons the valence band and the conduction band are the only energy bands that contribute to current flows in a semiconductor.

1.3.1 Energy Band Diagram

Figure 1–10 is the **energy band diagram** of a semiconductor, a small portion of Fig. 1–9. It shows the top edge of the valence band, denoted by E_v, and the bottom edge of the conduction band, denoted by E_c. The difference between E_c and E_v is the **band-gap energy** or **energy gap**, E_g. Clearly, $E_g = E_c - E_v$. For silicon, the energy gap is about 1.1 eV. The electrons in the valence band are those associated with the
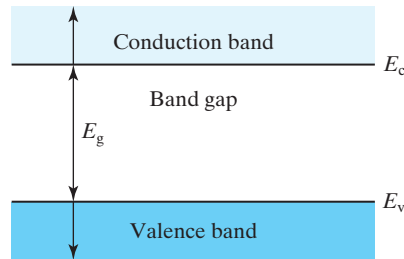
FIGURE 1–10 The energy band diagram of a semiconductor.

● **Photoconductor as Light Detector** ●

When light is absorbed by a semiconductor sample and electron–hole pairs are created as shown in Fig. 1–11, the number of electrons and holes (and therefore the conductivity of the semiconductor) increase in proportion to the light intensity. By putting two electrodes on the semiconductor and applying a voltage between the electrodes, one can measure the change in the semiconductor conductance and thus detect changes in light intensity. This simple yet practical photodetector is called a **photoconductor**.
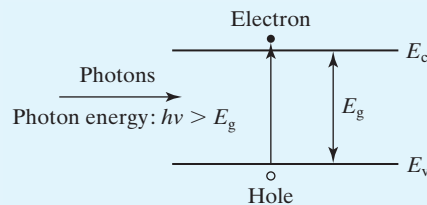

FIGURE 1–11 E_g can be determined from the minimum energy ($h\nu$) of photons that are absorbed by the semiconductor.

covalent bonds in the bond model discussed in the previous section, and the electrons in the conduction band are the conduction or mobile electrons. The band-gap energies of some semiconductors are listed in Table 1–1 to illustrate their wide range. The band-gap energy has strong influence on the characteristics and performance of optoelectronic devices (see Section 4.12.4 and Table 4–1). By mixing multiple semiconductors, the band-gap energy can be precisely tuned to desired values. This is widely practiced for optical semiconductor devices (see Section 4.13).

The band-gap energy can be determined by measuring the absorption of light by the semiconductor as a function of the photon energy, $h\nu$. The light is strongly absorbed only when $h\nu$ is larger than E_g. The absorbed photon energy is consumed to create an electron–hole pair as shown in Fig. 1–12. As $h\nu$ is reduced below E_g, the specimen becomes transparent to the light. E_g can be determined by observing this critical $h\nu$. Values of E_g listed in Table 1–1 are basically obtained in this way.

TABLE 1–1 • Band-gap energies of selected semiconductors.

Semiconductor	InSb	Ge	Si	GaAs	GaP	ZnSe	Diamond
E_g (eV)	0.18	0.67	1.12	1.42	2.25	2.7	6.0

EXAMPLE 1–1   Measuring the Band-Gap Energy

If a semiconductor is transparent to light with a wavelength longer than 0.87 μm, what is its band-gap energy?

SOLUTION:

Photon energy of light with 0.87 μm wavelength is, with c being the speed of light

$$h\nu = h\frac{c}{\lambda} = \frac{6.63 \times 10^{-34}\,(\text{J}\cdot\text{s}) \times 3 \times 10^{8}\,\text{m/s}}{0.87\ \mu\text{m}} = \frac{1.99 \times 10^{-19}\,(\text{J}\cdot\mu\text{m})}{0.87\ \mu\text{m}}$$

$$= \frac{1.99 \times 10^{-19}\,(\text{eV}\cdot\mu\text{m})}{1.6 \times 10^{-19} \times 0.87\ \mu\text{m}} = \frac{1.24\ (\text{eV}\cdot\mu\text{m})}{0.87\ \mu\text{m}} = 1.42\ \text{eV}$$

Therefore, the band gap of the semiconductor is 1.42 eV. The semiconductor is perhaps GaAs (see Table 1–1).

USEFUL RELATIONSHIP: $h\nu\ (\text{eV}) = \dfrac{1.24}{\lambda\ (\mu\text{m})}$

The visible spectrum is between 0.5 and 0.7 μm. (Silicon and GaAs have band gaps corresponding to the $h\nu$ of infrared light. Therefore they absorb visible light strongly and are opaque.) Some semiconductors such as indium and tin oxides have sufficiently large E_g's to be transparent to the visible light and be used as the transparent electrode in LCD (liquid crystal display) flat panel displays.
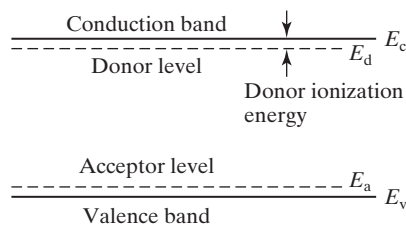

FIGURE 1–12 Energy levels of donors and acceptors.

1.3.2 Donors and Acceptors in the Band Model

The concept of donors and acceptors is expressed in the energy band model in the following manner. Although less important than E_c and E_v, two other energy levels are present in the energy band diagram, the donor energy level E_d and the acceptor energy level E_a (Fig. 1–12). Recall that it takes the donor ionization energy (about 50 meV) to free the extra electron from the donor atom into a conduction electron. Therefore, the donor electron, before the electron is freed, must occupy a state at about 50 meV below E_c. That is to say, $E_c - E_d$ is the donor ionization energy. Similarly, $E_a - E_v$ is the acceptor ionization energy (i.e., the

TABLE 1–2 ● Ionization energy of selected donors and acceptors in silicon.

	Donors			Acceptors		
Dopant	Sb	P	As	B	Al	In
Ionization energy, $E_c - E_d$ or $E_a - E_v$ (meV)	39	44	54	45	57	160

energy it takes for an acceptor atom to receive an extra electron from the valence band, creating a hole there). Some donor and acceptor ionization energies in silicon are listed in Table 1–2 for reference. As, P, Sb, and B are the most commonly used dopants for silicon. Acceptor and donor levels with small ionization energies, such as these four, are called **shallow levels**. Deep levels can be created with other impurities such as copper and gold, and they affect silicon properties in very different ways (see Section 2.6).

1.4 ● SEMICONDUCTORS, INSULATORS, AND CONDUCTORS ●

Based on the energy band model, we can now understand the differences among semiconductors, insulators, and conductors. A semiconductor has a nearly filled valence band and a nearly empty conduction band separated by a band gap as illustrated in Fig. 1–13a. The band diagram of an insulator is similar to that of a semiconductor except for a larger E_g, which separates a completely filled band and a completely empty band (see Fig. 1–13b). Totally filled bands and totally empty bands do not contribute to current conduction, just as there can be no motion of liquid in totally filled jars and totally empty jars. A conductor has a quite different energy band diagram. As depicted in Fig. 1–13c, a conductor has a partially filled band. This is the conduction band of the conductor and it holds the conduction electrons. The abundance of the conduction electrons makes the resistivity of a typical conductor much lower than that of semiconductors and insulators.

Why do some materials have a partially filled band and are therefore conductors? Each energy band can hold two electrons per atom.[1] This is why

(a) Si, semiconductor   (b) SiO$_2$, insulator   (c) Conductor
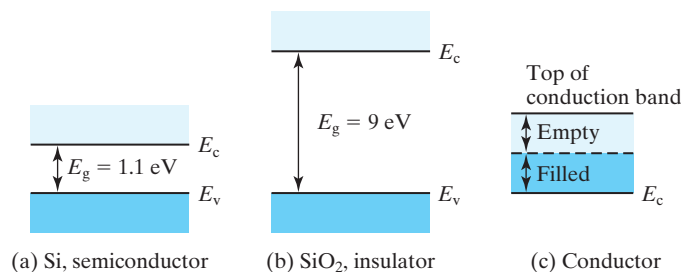
FIGURE 1–13 Energy band diagrams for a semiconductor (a), an insulator (b), and a conductor (c).

---

[1] This is a simplified picture. Actually, each band can hold two electrons per primitive cell, which may contain several atoms. A primitive cell is the smallest repeating structure that makes up a crystal. The darkened part of Fig. 1–2 is the primitive cell of the Si crystal.

elemental solids with odd atomic numbers (and therefore odd numbers of electrons) such as Au, Al, and Ag are conductors. Elements with even atomic numbers such as Zn and Pb can still be conductors because a filled band and an empty band may overlap in energy, thus leaving the combined band partially filled. These elements are known as **semimetals**.

An insulator has a filled valence band and an empty conduction band that are separated by a larger E_g. How large an E_g is large enough for the material to be classified as an insulator? There is no clear boundary, although 4 eV would be an acceptable answer. However, even diamond, with $E_g \sim 6$ eV (often cited as a textbook example of an insulator) exhibits semiconductor characteristics. It can be doped N type and P type, and electronic devices such as rectifiers and transistors have been made with diamond.

One may say that semiconductors differ from insulators in that semiconductors can be made N type or P type with low resistivities through impurity doping. This characteristic of the semiconductors is very important for device applications.

1.5 ● ELECTRONS AND HOLES ●

Although the term **electrons** can be used in conjunction with the valence band as in "nearly all the energy states in the valence band are filled with electrons," we should assume that the term usually means **conduction-band** electrons. **Holes** are the electron voids in the valence band. Electrons and holes carry negative and positive charge ($\pm q$) respectively. As shown in Fig. 1–14, a higher position in the energy band diagram represents a higher electron energy. The minimum conduction electron energy is E_c. Any energy above E_c is the electron kinetic energy. Electrons may gain energy by getting accelerated in an electric field and may lose energy through collisions with imperfections in the crystal.

A lower location in the energy diagram represents a higher hole energy as shown in Fig. 1–14. It requires energy to move a hole "downward" because that is equivalent to moving an electron upward. E_v is the minimum hole energy. *We may think of holes as bubbles in liquid, floating up in the energy band. Similarly, one may think of electrons as water drops that tend to fall to the lowest energy states in the energy band.*

1.5.1 Effective Mass

When an electric field, $\mathcal{E}$, is applied, an electron or a hole will accelerate according to

$$\text{Acceleration} = \frac{-q\mathcal{E}}{m_n} \quad \text{electrons} \tag{1.5.1}$$

$$\text{Acceleration} = \frac{q\mathcal{E}}{m_p} \quad \text{holes} \tag{1.5.2}$$

In order to describe the motion of electrons and holes with the laws of motion of the classical particles, we must assign **effective masses** (m_n and m_p) to them. The electron and hole effective masses of a few semiconductors are listed in Table 1–3.
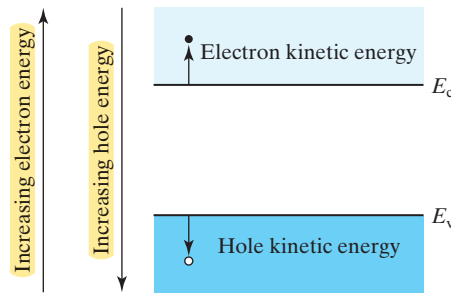
FIGURE 1–14 Both electrons and holes tend to seek their lowest energy positions. Electrons tend to fall in the energy band diagram. Holes float up like bubbles in water.

TABLE 1–3 ● Electron and hole effective masses, m_n and m_p, normalized to the free electron mass.

	Si	Ge	GaAs	InAs	AlAs
m_n/m_0	0.26	0.12	0.068	0.023	2.0
m_p/m_0	0.39	0.30	0.50	0.30	0.3

The electrons and holes in a crystal interact with a periodic coulombic field in the crystal. They surf over the periodic potential of the crystal, and therefore m_n and m_p are not the same as the free electron mass.

A complete description of the electrons in a crystal must be based on their wave characteristics, not just the particle characteristics. The electron wave function is the solution of the three-dimensional Schrödinger wave equation [2].

$$-\frac{\hbar}{2m_0}\nabla^2\psi + V(\mathbf{r})\psi = E\psi \tag{1.5.3}$$

where $\hbar = h/2\pi$ is the reduced Planck constant, m_0 is the free electron mass, $V(\mathbf{r})$ is the potential energy field that the crystal presents to the electron in the three-dimensional space, and E is the energy of the electron. The solution is of the form $\exp(\pm\mathbf{k}\cdot\mathbf{r})$, which represents an electron wave $\mathbf{k}$, called the **wave vector**, is equal at 2π/electron wavelength and is a function of E. In other words, for each $\mathbf{k}$ there is a corresponding E (see Fig. 4–27 for a schematic E–$\mathbf{k}$ diagram). It can further be shown [2] that, assuming the E–$\mathbf{k}$ relationship has spherical symmetry, an electric field, $\mathscr{E}$, would accelerate an electron wave packet with

$$\text{Acceleration} = -\frac{q\mathscr{E}}{\hbar^2}\frac{d^2E}{d\mathbf{k}^2} \tag{1.5.4}$$

In order to interpret the acceleration in the form of F/m, it is convenient to introduce the concept of the effective mass

$$\text{Effective mass} \equiv \frac{\hbar^2}{d^2E/d\mathbf{k}^2} \tag{1.5.5}$$

Each semiconductor material has a unique E–k relationship (due to the unique $V(r)$) for its conduction band and another unique E–k relationship for its valence band. Therefore, each semiconductor material has its unique m_n and m_p.

The values listed in Table 1–3 are experimentally measured values. These values agree well with the effective masses obtained by solving the Schrödinger wave equation with computers.

1.5.2 How to Measure the Effective Mass[2]

If you wonder how one may measure the effective mass of electrons or holes in a semiconductor, let us study a powerful technique called **cyclotron resonance**.

Consider an electron in an N-type semiconductor located in a magnetic field, B, as shown in Fig. 1–15. A moving electron will trace out a circular path in a plane normal to B. (In addition, there may be electron motion parallel to B. We may ignore this velocity component for the present analysis). The magnetic field exerts a Lorentzian force of qvB, where v is the electron velocity and B is the magnetic flux density. Equating this force to the centripetal force corresponding to the circular motion with radius r, we obtain

$$\frac{m_n v^2}{r} = qvB \tag{1.5.6}$$

$$v = \frac{qBr}{m_n} \tag{1.5.7}$$

The frequency of the circular motion is

$$f_{cr} = \frac{v}{2\pi r} = \frac{qB}{2\pi m_n} \tag{1.5.8}$$

This is the cyclotron resonance frequency. Notice that the resonance frequency is independent of r and v. Now, if a circularly polarized electric field of the same frequency f_{cr} (typically in the gigahertz range) is applied to this semiconductor, the
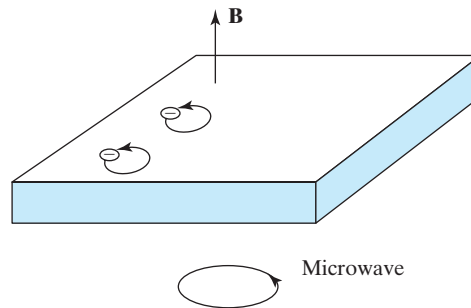

FIGURE 1–15 The motion of electrons in an N-type semiconductor in the presence of a magnetic field, **B**, and a microwave with rotating electric field (the direction of rotation is indicated by the arrow).

---

[2] This section may be omitted in an accelerated course.

electrons will strongly absorb the microwave energy. They do so by accelerating to a higher velocity and tracing circles of increasing radius [see Eq. (1.5.7)] without changing their frequency of circular motion [see Eq. (1.5.8)], losing the energies through collisions, and starting the acceleration process again. Obviously, the absorption would be much weaker if the frequency of the applied field is not equal to f_{cr}, i.e., when the applied field is out of sync with the electron motion.

By varying the frequency of the electric field or varying **B** until a peak in absorption is observed, one can calculate m_n using Eq. (1.5.8). One can also perform the same measurement on a P-type semiconductor to measure the effective mass of holes.

1.6 ● DENSITY OF STATES ●

It is useful to think of an energy band as a collection of discrete energy states. Figure 1–16a emphasizes this picture. In quantum mechanics terms, each state represents a unique spin (up and down) and unique solution to the Schrodinger's wave equation for the periodic electric potential function of the semiconductor [3]. Each state can hold either one electron or none. If we count the number of states in a small range of energy, ΔE, in the conduction band, we can find the **density of states**:

$$D_c(E) \equiv \frac{\text{number of states in } \Delta E}{\Delta E \times \text{volume}} \tag{1.6.1}$$

This conduction-band density of states is a function of E (i.e., where ΔE is located). Similarly, there is a valence-band density of states, $D_v(E)$. D_c and D_v, graphically illustrated in Fig. 1–16b, can be shown to be proportional to $\sqrt{E - E_c}$ and $\sqrt{E_v - E}$ at least for a range of E. The derivation is presented in Appendix I, "Derivation of the Density of States."

$$D_c(E) = \frac{8\pi m_n \sqrt{2m_n(E - E_c)}}{h^3}, \quad E \geq E_c \tag{1.6.2a}$$

$$D_v(E) = \frac{8\pi m_p \sqrt{2m_p(E_v - E)}}{h^3}, \quad E \leq E_v \tag{1.6.2b}$$
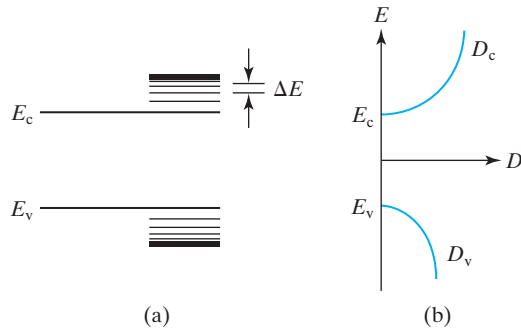

FIGURE 1–16 (a) Energy band as a collection of discrete energy states. (b) D is the density of the energy states.

$D_c(E)$ and $D_v(E)$ have the dimensions of number per cubic centimeter per electronvolt.

It follows from Eq. (1.6.1) that the product $D_c(E) \, dE$ and $D_v(E) \, dE$ are the numbers of energy states located in the energy range between E and $E + dE$ per cubic centimeter of the semiconductor volume. We will use this concept in Section 1.8.1.

1.7   •   THERMAL EQUILIBRIUM AND THE FERMI FUNCTION   •

We have mentioned in Section 1.6 that most of the conduction-band electrons will be found near E_c, where the electron energy is the lowest. Most holes will "float" toward E_v, where the hole energy is the lowest. In this section, we will examine the distribution of electrons and holes in greater detail.

1.7.1 An Analogy for Thermal Equilibrium

Let us perform the following mental experiment. Spread a thin layer of sand on the bottom of a shallow dish sitting on a table as shown in Fig. 1–17. The sand particles represent the electrons in the conduction band. A machine shakes the table and therefore the dish up and down. The vibration of the dish represents the thermal agitation experienced by the atoms, electrons and holes at any temperature above absolute zero. The sand will move and shift until a more or less level surface is created. This is the equilibrium condition. The **equilibrium** condition is the lowest energy configuration in the *presence of the thermal agitation*. If a small sand dune exists as a protrusion above the flat surface, it would not be the lowest energy condition because the gravitational energy of the system can be reduced by flattening the dune. This is the equivalent of electrons preferring to occupy the lower energy states of the conduction band.

Notice, however, that the agitation prevents the particles from taking only the lowest possible energy positions. This fact can be dramatized by vibrating the dish more vigorously. Now, some sand particles jump up into the air and fall back. The system is at the lowest possible energy consistent with the presence of the "thermal" agitation. This is **thermal equilibrium**.

Because the sand particles bounce up and down, some higher energy states, in the air, are occupied, too. There is a lower probability (smaller fraction of time) for the sand particles to be higher in the air, i.e., to occupy higher energy states. Similarly, electrons and holes in semiconductors receive and exchange energy from
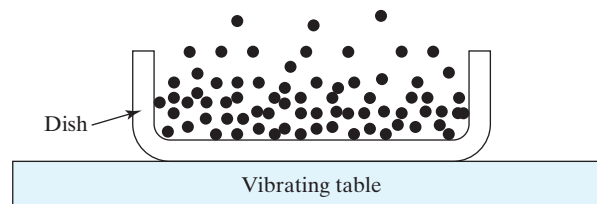
Dish

Vibrating table

FIGURE 1–17 Elevations of the sand particles in the dish on a vibrating table represent the energies of the electrons in the conduction band under the agitation of thermal energy.

or with the crystal and one another and *every energy state in the conduction and valence bands has a certain probability of being occupied by an electron*. This probability is quantified in the next section.

1.7.2 Fermi Function—The Probability of an Energy State Being Occupied by an Electron

One can derive an expression to describe the probability of finding an electron at a certain energy. A statistical thermodynamic analysis without regard to the specifics of how particles bounce off the atoms or one another and only assuming that the number of particles and the total system energy are held constant yield the following result. (The derivation is presented in Appendix II, "Derivation of the Fermi–Dirac Distribution Function.)

$$f(E) = \frac{1}{1 + e^{(E - E_F)/kT}} \qquad (1.7.1)$$

An equation that is highlighted with a box, such as Eq. (1.7.1), is particularly important and often cited. Equation (1.7.1) is the **Fermi function**, or the **Fermi–Dirac distribution function**, or the **Fermi–Dirac statistics**. E_F is called the **Fermi energy** or the **Fermi level**. $f(E)$ is the probability of a state at energy E being occupied by an electron. Figure 1–18 depicts the Fermi function. At large E (i.e., $E - E_F \gg kT$) the probability of a state being occupied decreases exponentially with increasing E. In this energy region, Eq. (1.7.1) can be approximated with

$$f(E) \approx e^{-(E - E_F)/kT} \qquad (1.7.2)$$

Equation (1.7.2) is known as the **Boltzmann approximation**. In the low energy region (i.e., $E - E_F \ll -kT$), the occupation probability approaches 1. In other words, the low energy states tend to be fully occupied. Here Eq. (1.7.1) can be approximated as

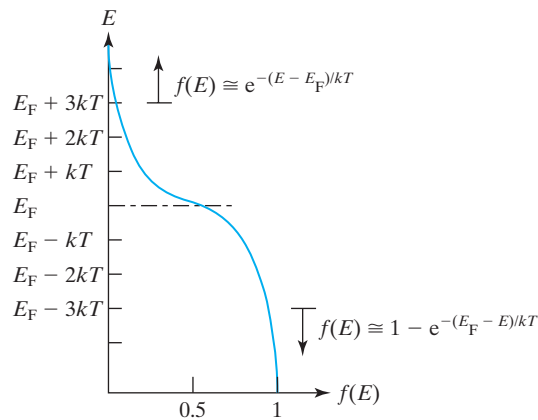$$f(E) \approx 1 - e^{-(E_F - E)/kT} \qquad (1.7.3)$$

FIGURE 1–18 The Fermi function diagram. For an interactive illustration of the Fermi function, see http://jas.eng.buffalo.edu/education/semicon/fermi/functionAndStates/functionAndStates.html

In this energy region, the probability of a state *not* being occupied, i.e., being *occupied by a hole* is

$$1 - f(E) \approx e^{-(E_F - E)/kT} \qquad (1.7.4)$$

From Eq. (1.7.1) and Fig. 1–18, the probability of occupation at E_F is 1/2. The probability approaches unity if E is much lower than E_F, and approaches zero at E much higher than E_F. *A very important fact to remember about E_F is that* there is only one Fermi level in a system at thermal equilibrium. This fact will be used often in the rest of this book. Later, you will see that the value of E_F depends on the number of electrons or holes present in the system.

● **What Determines E_F ?** ●

This example is designed to show that the Fermi level, E_F, is determined by the available electrons and states in the system. Figure 1–19 shows the energy states of an electron system at room temperature. Each energy state can hold either one electron or none, i.e., be either occupied or empty. It is known that there is only one electron in this system. Since there is only one electron in the system, the sum of the probabilities that each state is occupied must be equal to 1. By trial and error, convince yourself that only one specific E_F, somewhere above E_1 and below E_2, can satisfy this condition. Do not calculate this value. *Hint*: Ask yourself how many electrons will be in the system if E_F is, say, above E_3 and how many if E_F is below E_1.

$$
\begin{array}{ll}
\rule{3cm}{0.4pt} & E_3 = 0 \text{ eV} \\[2em]
\rule{3cm}{0.4pt} & E_2 = -1 \text{ eV} \\[5em]
E_F \; \text{--·--·--·--} \\
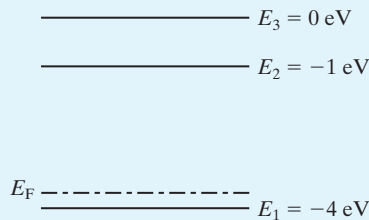\rule{3cm}{0.4pt} & E_1 = -4 \text{ eV}
\end{array}
$$

FIGURE 1–19 A simple electron system at room temperature for illustration of what determines the Fermi energy, E_F.

EXAMPLE 1–2   Oxygen Concentration versus Altitude

We all know that there is less oxygen in the air at higher altitudes. What is the ratio of the oxygen concentration at 10 km above sea level, N_h, to the concentration at sea level, N_0, assuming a constant temperature of 0°C?

SOLUTION:

There are fewer oxygen molecules at higher altitudes because the gravitational potential energy of an oxygen molecule at the higher altitude, E_h, is larger than at sea level, E_0. According to Eq. (1.7.2)

$$\frac{N_h}{N_0} = \frac{e^{-E_h/kT}}{e^{-E_0/kT}} = e^{-(E_h - E_0)/kT}$$

$E_0 - E_h$ is the potential energy difference, i.e., the energy needed to lift an oxygen molecule from sea level to 10 km.

$$E_h - E_0 = \text{altitude} \times \text{weight of O}_2 \text{ molecule} \times \text{acceleration of gravity}$$

$$= 10^4 m \times \text{O}_2 \text{ molecular weight} \times \text{atomic mass unit} \times 9.8 \ m \cdot s^{-2}$$

$$= 10^4 m \times 32 \times 1.66 \times 10^{-27} kg \times 9.8 \ m \cdot s^{-2}$$

$$= 5.2 \times 10^{-21} J$$

$$\therefore \frac{N_h}{N_0} = e^{-5.2 \times 10^{-21} J / 1.38 \times 10^{-10} J \cdot K^{-1} \times 273 K}$$

$$= e^{-1.38} = 0.25$$

So, the oxygen concentration at 10 km is 25% of the sea level concentration.

This example and the sand-in-a-dish analogy are presented to demystify the concept of equilibrium, and to emphasize that each electron energy state has a probability of being occupied that is governed by the Fermi function.

Additional question: See Problem 1.4 for a follow-up question.

1.8 ● ELECTRON AND HOLE CONCENTRATIONS ●

We have stated that if a semiconductor is doped with 10^{16} donors per cubic centimeter, the electron concentration will be 10^{16}cm^{-3}. But, what would the hole concentration be? What are the carrier concentrations in undoped semiconductors? These questions will be answered after the important relationships between the Fermi level and the carrier concentrations are derived in this section.

1.8.1 Derivation of *n* and *p* from *D(E)* and *f(E)*

First, we will derive the concentration of electrons in the conduction band, known as the **electron concentration**. Since $D_c(E) \ dE$ is the number of energy states between E and $E + dE$ for each cubic centimeter, the product $f(E)D_c(E) \ dE$ is then the number of electrons between E and $E + dE$ per cubic centimeter of the semiconductor. Therefore, the number of electrons per cubic centimeter in the entire conduction band is

$$n = \int_{E_c}^{\text{Top of conduction band}} f(E)D_c(E) \ dE \tag{1.8.1}$$

Graphically, this integration gives the shaded area in Fig. 1–20, which yields the density of electrons. We now substitute Eqs. (1.6.2a) and (1.7.2) into Eq. (1.8.1) and set the upper limit of integration at infinity. Resetting the upper limit is acceptable because of the rapid fall of $f(E)$ with increasing E as shown in Fig. 1–20. This allows us to obtain a closed form expression for *n*
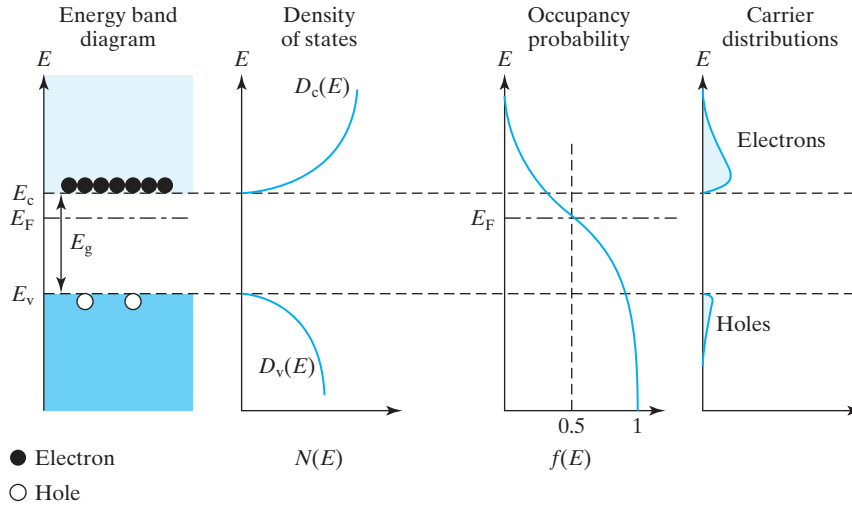
FIGURE 1–20 Schematic band diagram, density of states, Fermi–Dirac distribution, and carrier distributions versus energy.

$$n = \frac{8\pi m_n \sqrt{2m_n}}{h^3} \int_{E_c}^{\infty} \sqrt{E - E_c} \ e^{-(E - E_F)/kT} dE \tag{1.8.2}$$

$$= \frac{8\pi m_n \sqrt{2m_n}}{h^3} e^{-(E_c - E_F)/kT} \int_0^{\infty} \sqrt{E - E_c} \ e^{-(E - E_c)/kT} d(E - E_c) \tag{1.8.3}$$

Introducing a new variable

$$x = (E - E_c)/kT$$

reveals that the integral in Eq. (1.8.3) is of a form known as a **gamma function** and is equal to $\sqrt{\pi}/2$, i.e.,

$$\int_0^{\infty} \sqrt{x} \ e^{-x} dx = \sqrt{\pi}/2 \tag{1.8.4}$$

Applying Eq. (1.8.4) to Eq. (1.8.3) leads to the following two equations:

$$n = N_c \, e^{-(E_c - E_F)/kT} \tag{1.8.5}$$

$$N_c \equiv 2 \left[\frac{2\pi m_n kT}{h^2} \right]^{3/2} \tag{1.8.6}$$

N_c is called the **effective density of states**. Equation (1.8.5) is an important equation and should be memorized. It is easy to remember this equation if we understand why N_c is called the **effective** density of states. It is as if all the energy states in the

TABLE 1–4 ● Values of N_c and N_v for Ge, Si, and GaAs at 300 K.

	Ge	Si	GaAs
N_c (cm^{-3})	1.04×10^{19}	2.8×10^{19}	4.7×10^{17}
N_v (cm^{-3})	6.0×10^{18}	1.04×10^{19}	7.0×10^{18}

conduction band were *effectively* squeezed into a single energy level, E_c, which can hold N_c electrons (per cubic centimeter). As a result, the electron concentration in Eq. (1.8.5) is simply the product of N_c and the probability that an energy state at E_c is occupied.

An expression for the hole concentration can be derived in the same way. The probability of an energy state being occupied by a hole is the probability of it *not* being occupied by an electron, i.e., $1 - f(E)$. Therefore,

$$p = \int_{\substack{\text{Valence band} \\ \text{bottom}}}^{E_v} D_v(E)(1 - f(E))\, dE \tag{1.8.7}$$

Substituting Eqs. (1.6.2b) and (1.7.4) into Eq. (1.8.7) yields

$$p = N_v\, e^{-(E_F - E_v)/kT} \tag{1.8.8}$$

$$N_v \equiv 2\left[\frac{2\pi m_p kT}{h^2}\right]^{3/2} \tag{1.8.9}$$

N_v is the **effective density of states of the valence band**. (The full name of N_c is the effective density of states of the conduction band.) The values of N_c and N_v, both about 10^{19}cm^{-3}, differ only because m_n and m_p are different. N_c and N_v vary somewhat from one semiconductor to another because of the variation in the effective masses, too. N_c and N_v for Ge, Si, and GaAs are listed in Table 1–4.[3]

1.8.2 Fermi Level and the Carrier Concentrations

We will use Eqs. (1.8.5) and (1.8.8) time and again. Right now, they can help to remove the mystery of the Fermi level by linking E_F to the electron and hole concentrations.

Figure 1–21 shows the location of the Fermi level as a function of the carrier concentration. Note that the solid lines stop when E_F is about 20 meV ($\sim kT$) from

---

[3] The effective mass in Table 1–3 is called the *conductivity effective mass* and is an average over quantum mechanical wave vectors appropriate for describing carrier motions [3]. The effective mass in Eqs. (1.6.2a) and (1.6.2b), (1.8.6), and (1.8.9) is called the *density-of-states effective mass* and has a somewhat different value (because it is the result of a different way of averaging that is appropriate for describing the density of states).
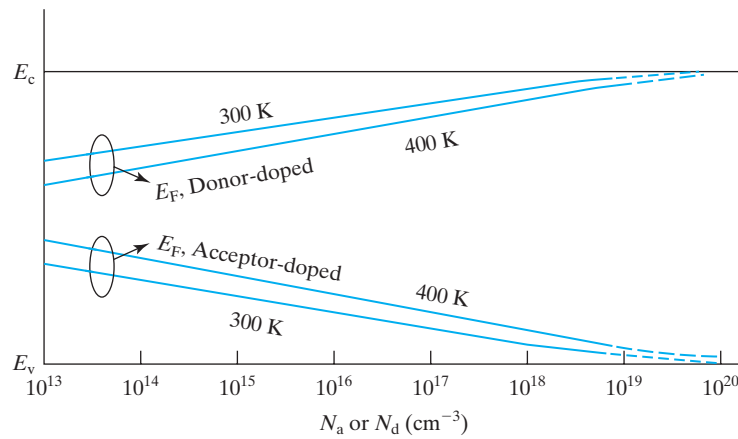
FIGURE 1–21 Location of Fermi level vs. dopant concentration in Si at 300 and 400 K.

EXAMPLE 1–3 **Finding the Fermi Level in Si**

Where is E_F located in the energy band of silicon, at 300K with $n = 10^{17}\text{cm}^{-3}$?
And for $p = 10^{14}\text{cm}^{-3}$?

SOLUTION: From Eq. (1.8.5)

$$E_c - E_F = kT \cdot \ln(N_c/n)$$

$$= 0.026\ln(2.8 \times 10^{19}/10^{17})$$

$$= 0.146 \text{ eV}$$

Therefore, E_F is located at 146 meV below E_c, as shown in Fig. 1–22a.

For $p = 10^{14}\text{cm}^{-3}$, from Eq. (1.8.8),

$$E_F - E_v = kT \cdot \ln(N_v/p)$$

$$= 0.026\ln(1.04 \times 10^{19}/10^{14})$$

$$= 0.31 \text{ eV}$$

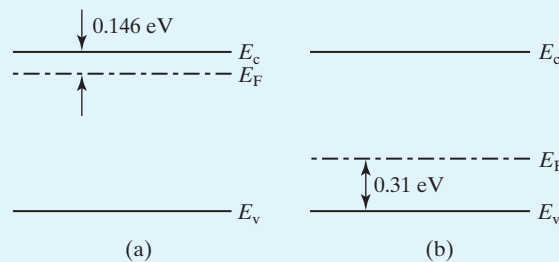Therefore E_F is located at 0.31 eV above E_v.

FIGURE 1–22 Location of E_F when $n = 10^{17}\text{cm}^{-3}$ (a), and $p = 10^{14}\text{cm}^{-3}$ (b).

E_c or E_v. Beyond this range, the use of the Boltzmann approximation in the derivation of Eqs. (1.8.5) and (1.8.8) is not quantitatively valid. Those equations are not accurate when the semiconductor is heavily doped ($>10^{19} \text{cm}^{-3}$) or **degenerate**. *Please remember that n decreases as E_F moves farther below E_c, and vice versa; p decreases as E_F moves farther above E_v.*

1.8.3 The *np* Product and the Intrinsic Carrier Concentration

Since E_F cannot be close to both E_c and E_v, n and p cannot both be large numbers at the same time. When Eqs. (1.8.5) and (1.8.8) are multiplied together, we obtain

$$np = N_c N_v e^{-(E_c - E_v)/kT} = N_c N_v e^{-E_g/kT} \tag{1.8.10}$$

Equation (1.8.10) states that the *np* product is a constant for a given semiconductor and T, independent of the dopant concentrations. It is an important relationship and is usually expressed in the following form:

$$np = n_i^2 \tag{1.8.11}$$

$$n_i = \sqrt{N_c N_v}\, e^{-E_g/2kT} \tag{1.8.12}$$

According to Eq. (1.8.11), there are always some electrons and holes present—whether dopants are present or not. If there are no dopants present, the semiconductor is said to be **intrinsic**. In an intrinsic semiconductor, the nonzero n and p are the results of thermal excitation, which moves some electrons from the valence band into the conduction band. Since such movements create electrons and holes in pairs, $n = p$ in intrinsic semiconductors. This fact and Eq. (1.8.11) immediately suggest that, in intrinsic semiconductors,

$$n = p = n_i \tag{1.8.13}$$

Therefore, n_i is called the **intrinsic carrier concentration**. n_i is a strong function of E_g and T according to Eq. (1.8.12), but is independent of the dopant concentration. n_i at room temperature is roughly 10^{10}cm^{-3} for Si and 10^7cm^{-3} for GaAs, which has a larger band gap than Si. For silicon, the np product is therefore 10^{20}cm^{-6} regardless of the conductivity type (P type or N type) and the dopant concentrations.

One may explain why the *np* product is a constant this way: the electron–hole recombination rate is proportional to the np product. When $np = n_i^2$, the recombination rate happens to be equal to the rate of thermal generation of electron–hole pairs. This is the same mass action principle that keeps the product of the concentrations of [H$^+$] and [OH$^-$] constant in aqueous solutions, whether strongly acidic, strongly alkaline, or neutral.

EXAMPLE 1–4   **Carrier Concentrations**

QUESTION: What is the hole concentration in an N-type semiconductor with 10^{15}cm^{-3} of donors?

SOLUTION: For each ionized donor, an electron is created. Therefore $n = 10^{15}\text{cm}^{-3}$.

$$p = \frac{n_i^2}{n} \approx \frac{10^{20}\text{ cm}^{-3}}{10^{15}\text{ cm}^{-3}} = 10^5\text{cm}^{-3}$$

With a modest temperature increase of 60°C, n remains the same at 10^{15}cm^{-3}, while p increases by about a factor of 2300 because n_i^2 increases according to Eq. (1.8.12).

QUESTION: What is n if $p = 10^{17}\text{cm}^{-3}$ in a P-type silicon wafer?

SOLUTION: $n = \dfrac{n_i^2}{p} = \dfrac{10^{20}}{10^{17}} = 10^3\text{cm}^{-3}$

The electron and hole concentrations in a semiconductor are usually very different. In an N-type semiconductor, the abundant electrons are called the **majority carriers** and the almost nonexistent holes are the **minority carriers**. In P-type semiconductors, the holes are the majority carriers and the electrons are the minority carriers.

EXAMPLE 1–5   **The Intrinsic Fermi Level**

In an intrinsic semiconductor, $n = p$. Therefore $E_c - E_F \approx E_F - E_v$ and the Fermi level is nearly at the middle of the band gap, i.e., $E_F \approx E_c - E_g/2$. This level is called the **intrinsic Fermi level**, E_i. Here we derive a more exact expression for E_i. Rewriting Eq. (1.8.12) for $\ln n_i$, yields

$$\ln n_i = \ln \sqrt{N_c N_v} - E_g/2kT$$

Writing Eq. (1.8.5) for the intrinsic condition where $n = n_i$,

$$n_i = N_c e^{-(E_c - E_i)/kT} \tag{1.8.14}$$

$$\therefore E_i = E_c - kT\ln\frac{N_c}{n_i} = E_c + kT\ln n_i - kT\ln N_c = E_c - \frac{E_g}{2} - kT\ln\sqrt{\frac{N_c}{N_v}} \tag{1.8.15}$$

We see that E_i would be at the midgap, $E_c - E_g/2$, if $N_c = N_v$. For silicon, E_i is very close to the midgap and the small last term in Eq. (1.8.15) is only of academic interest.

1.9 ● GENERAL THEORY OF *n* AND *p* ●

The shallow donor and acceptor levels (E_d and E_a) in Fig. 1–12 are energy states and their occupancy by electrons is governed by the Fermi function. Since E_d is usually a few kT above E_F, the donor level is nearly empty of electrons. We say that nearly all the donor atoms are ionized (have lost the extra electrons). Similarly, all the acceptor atoms are ionized. *For simplicity, we can assume that all shallow donors and acceptors are ionized.*

EXAMPLE 1–6 Complete Ionization of the Dopant Atoms

In a silicon sample doped with $10^{17} cm^{-3}$ of phosphorus atoms, what fraction of the donors are not ionized (i.e., what fraction are occupied by the "extra" electrons)?

SOLUTION:

First assume that all the donors are ionized and each donor donates an electron to the conduction band.

$$n = N_d = 10^{17} cm^{-3}$$

From Fig. 1–20, Example 1–3, E_F is located at 146 meV below E_c. The donor level E_d is located at 45 meV below E_c for phosphorus (see Table 1–2 and Figure 1–23).

The probability that a donor is not ionized, i.e., the probability that it is occupied by the "extra" electron, according to Eq. (1.7.1), is

$$\text{Probability of non-ionization} \approx \frac{1}{1 + \frac{1}{2} e^{(E_d - E_F)/kT}}$$

$$\frac{1}{1 + \frac{1}{2} e^{((146-45)meV)/26meV}} = 3.9\%$$

(The factor 1/2 in the denominators stems from the complication that a donor atom can hold an electron with upspin or downspin. This increases the probability that a donor state is occupied by an electron.)

Therefore, it is reasonable to assume complete ionization, i.e., $n = N_d$.
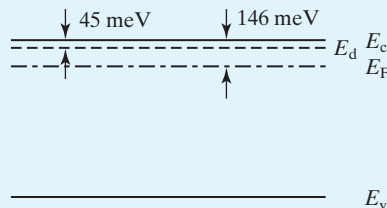

FIGURE 1–23 Location of E_F and E_d. Not to scale.

DISCUSSION: You may have noticed that as N_d increases, E_F rises toward E_d and the probability of nonionization can become quite large. In reality, the impurity level broadens into an **impurity band** that merges with the conduction band in heavily doped semiconductor (i.e., when donors or acceptors are close to one another). This happens for the same reason energy levels broaden into bands when atoms are brought close to one another to form a crystal (see Fig. 1–9). The electrons in the impurity band are also in the conduction band. *Therefore, the assumption of $n = N_d$ (or complete ionization) is reasonable even at very high doping densities. The same holds true in P-type materials.*

There are four types of charged species in a semiconductor: electrons, holes, positive donor ions, and negative acceptor ions. Their densities are represented by the symbols n, p, N_d, and N_a. In general, all samples are free of net charge. **Charge neutrality** requires that the densities of the negative particles and positive particles are equal:

$$n + N_a = p + N_d \tag{1.9.1}$$

Equations (1.8.11) and (1.9.1) can be solved for n and p:

$$n = \frac{N_d - N_a}{2} + \left[\left(\frac{N_d - N_a}{2}\right)^2 + n_i^2\right]^{1/2} \tag{1.9.2a}$$

$$p = \frac{N_a - N_d}{2} + \left[\left(\frac{N_a - N_d}{2}\right)^2 + n_i^2\right]^{1/2} \tag{1.9.2b}$$

Although it is interesting to know that n and p can be calculated for arbitrary values of N_a and N_d, the complicated Eq. (1.9.2) is rarely used. Instead, one of the following two cases is almost always valid:

1. $N_d - N_a \gg n_i$ (i.e., N type),

$$n = N_d - N_a \tag{1.9.3a}$$

$$p = n_i^2/n \tag{1.9.3b}$$

If, furthermore, $N_d \gg N_a$, then

$$n = N_d \quad \text{and} \quad p = n_i^2/N_d \tag{1.9.4}$$

2. $N_a - N_d \gg n_i$ (i.e., P type),

$$p = N_a - N_d \tag{1.9.5a}$$

$$n = n_i^2/p \tag{1.9.5b}$$

If, furthermore, $N_a \gg N_d$, then

$$p = N_a \quad \text{and} \quad n = n_i^2/N_a \tag{1.9.6}$$

We have intuitively assumed Eqs. (1.9.4) and (1.9.6) to be true in the previous sections. It is worthwhile to remember that Eqs. (1.9.3) and (1.9.5) as the more exact expressions. We see that an acceptor can effectively negate the effect of a donor in Eq. (1.9.3a) and vice versa in Eq. (1.9.5a). This fact is known as **dopant compensation**. One can even start with P-type Si and convert a portion of it into N-type simply by adding enough donors. This is one of the techniques employed to make complex devices.

EXAMPLE 1–7 **Dopant Compensation**

What are *n* and *p* in a Si sample with $N_d = 6 \times 10^{16} \text{cm}^{-3}$ and $N_a = 2 \times 10^{16} \text{cm}^{-3}$? With additional $6 \times 10^{16} \text{cm}^{-3}$ of acceptors?

SOLUTION: As shown in Fig. 1–24a:

$$n = N_d - N_a = 4 \times 10^{16} \text{cm}^{-3}$$

$$p = n_i^2/n = 10^{20}/4 \times 10^{16} = 2.5 \times 10^3 \text{cm}^{-3}$$

With the additional acceptors, $N_a = 2 \times 10^{16} + 6 \times 10^{16} = 8 \times 10^{16} \text{cm}^{-3}$, holes become the majority,

$$p = N_a - N_d = 8 \times 10^{16} - 6 \times 10^{16} = 2 \times 10^{16} \text{cm}^{-3}$$

$$n = n_i^2/p = 10^{20}/(2 \times 10^{16}) = 5 \times 10^3 \text{cm}^{-3}$$

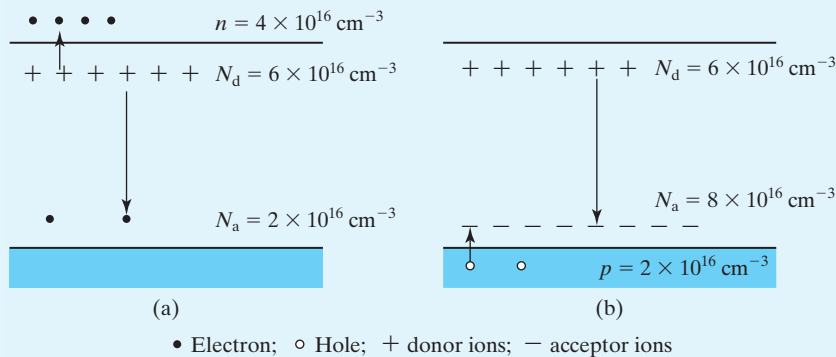The addition of acceptors has converted the Si to P-type as shown in Fig. 1–24b.

• Electron; ○ Hole; + donor ions; − acceptor ions

FIGURE 1–24 Graphical illustration of dopant compensation.

1.10 ● CARRIER CONCENTRATIONS AT EXTREMELY HIGH AND LOW TEMPERATURES[4] ●

At very high temperatures, n_i is large [see Eq. (1.8.12)], and it is possible to have $n_i \gg |N_d - N_a|$. In that case, Eq. (1.9.2) becomes

$$n = p = n_i \tag{1.10.1}$$

In other words, the semiconductor becomes "intrinsic" at very high temperatures.

At the other extreme of very low temperature, E_F may rise above E_d, and most of the donor (or acceptor, in the case of P-type material) atoms can remain nonionized. The fifth electrons stay with the donor. This phenomenon is called **freeze-out**. In this case, if the doping is not heavy enough to form an impurity band (see Section 1.9), the dopants are not totally ionized. The carrier concentration may be significantly less than the dopant concentration. The exact analysis is complicated, but the result is [4]

$$n = \left[\frac{N_c N_d}{2}\right]^{1/2} e^{-(E_c - E_d)/2kT} \tag{1.10.2}$$

Freeze-out is a concern when semiconductor devices are operated at, for example, the liquid–nitrogen temperature (77 K) in order to achieve low noise and high speed.

Figure 1–25 summarizes the temperature dependence of majority carrier concentrations. The slope of the curve in the intrinsic regime is $E_g/2k$, and the slope in the freeze-out portion is $(E_c - E_d)/2k$ (according to Eq. (1.10.2)). These facts may be used to determine E_g and $E_c - E_d$.
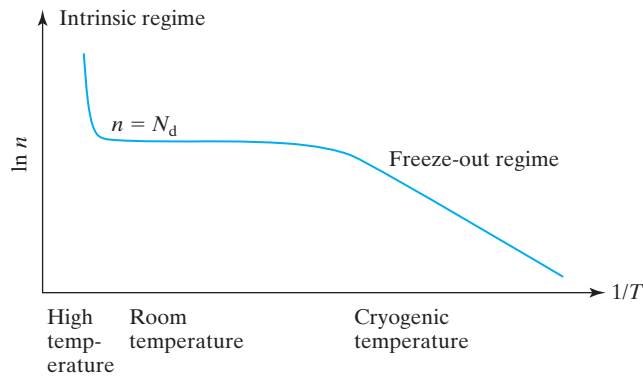

FIGURE 1–25 Variation of carrier concentration in an N-type semiconductor over a wide range of temperature.

___

[4] This section may be omitted in an accelerated course.

● **Infrared Detector Based on Freeze-Out** ●

Often it is desirable to detect or image the black-body radiation emitted by warm objects, e.g., to detect tumors (which restrict blood flow and produce cold spots), to identify inadequately insulated building windows, to detect people and vehicles at night, etc. This requires a photodetector that responds to photon energies around 0.1 eV. For this purpose, one can use a semiconductor photoconductor with E_g less than 0.1 eV, such as HgPbTe operating in the mode shown in Fig. 1–11. Alternatively, one can use a more common semiconductor such as doped Si operating in the freeze-out mode shown in Fig. 1–26. In Fig. 1–26, conduction electrons are created when the infrared photons provide the energy to ionize the donor atoms, which are otherwise frozen-out. The result is a lowering of the detector's electrical resistance, i.e., photoconductivity.

At long enough wavelength or low enough photon energy $h\nu$, light will no longer be absorbed by the specimen shown in Fig. 1–26. That critical $h\nu$ corresponds to $E_c - E_d$. This is a method of measuring the dopant ionization energy, $E_c - E_d$.
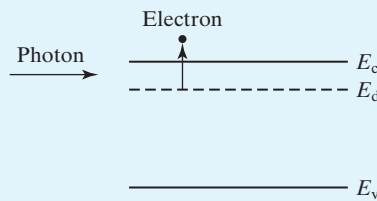

FIGURE 1–26 Infrared photons can ionize the frozen-out donors and produce conduction electrons.

1.11 ● CHAPTER SUMMARY ●

In a silicon crystal, each Si atom forms covalent bonds with its four neighbors. In an intrinsic Si crystal, there are few mobile electrons and holes. Their concentrations are equal to n_i ($\sim 10^{10}$ cm^{-3} for Si), the intrinsic carrier concentration. The bond model explains why group V atoms can serve as donors and introduce electrons, and group III atoms can serve as acceptors and introduce holes. Electrons and holes should be thought of as particles of equal importance but carrying negative and positive charge, respectively, and having effective masses, m_n and m_p, which are a fraction of the free electron mass.

The band model is needed for quantitative analysis of semiconductors and devices. The valence band and conduction band are separated by an energy gap. This band gap is 1.12 eV for Si. Thermal agitation gives each energy state a certain probability of being occupied by an electron. That probability is expressed by the Fermi function, which reduces to simple exponential functions of $(E - E_F)/kT$ for E more than a few kT's above the Fermi level, E_F. E_F is related to the density of electrons and holes in the following manner:

$$n = N_c\, e^{-(E_c - E_F)/kT} \tag{1.8.5}$$

$$p = N_v \, e^{-(E_F - E_v)/kT} \tag{1.8.8}$$

The effective densities of states, N_c and N_v, are around 10^{19}cm^{-3}. The majority carrier concentrations are

$$n = N_d - N_a \quad \text{for N-type semiconductor} \tag{1.9.3a}$$

$$p = N_a - N_d \quad \text{for P-type semiconductor} \tag{1.9.5a}$$

where N_d and N_a are the concentrations of donors and acceptors, which are usually assumed to be completely ionized. The minority carrier concentrations can be found from

$$np = n_i^2 \tag{1.8.11}$$

n_i is the intrinsic carrier concentration, about 10^{10}cm^{-3} for Si at 300 K. It is a function of E_g and T.

   You are now ready to study how electrons and holes move and produce current in the next chapter.

● PROBLEMS ●

● Visualization of the Silicon Crystal ●

1.1 **(a)** How many silicon atoms are there in each unit cell?

   (b) How many silicon atoms are there in one cubic centimeter?

   (c) Knowing that the length of a side of the unit cell (the silicon **lattice constant**) is 5.43 Å, Si atomic weight is 28.1, and the Avogdaro's number is 6.02×10^{23} atoms/mole, find the silicon density in g/cm^3.

● Fermi Function ●

1.2 **(a)** Under equilibrium condition, what is the probability of an electron state being occupied if it is located at the Fermi level?

   (b) If E_F is positioned at E_c, determine the probability of finding electrons in states at $E_c + kT$. (A numerical answer is required.)

   (c) The probability of a state being filled at $E_c + kT$ is equal to the probability of a state being empty at $E_c + 3\,kT$. Where is the Fermi level located?

1.3 **(a)** What is the probability of an electron state being filled if it is located at the Fermi level?

   (b) If the probability that a state being filled at the conduction band edge (E_c) is precisely equal to the probability that a state is empty at the valence band edge (E_v), where is the Fermi level located?

   (c) The Maxwell–Boltzmann distribution is often used to approximate the Fermi-Dirac distribution function. On the same set of axes, sketch both distributions as a function of $(E - E_F)/kT$. Consider only positive values of $E - E_F$. For what range of $(E - E_F)/kT$ is the Maxwell–Boltzmann approximation accurate to within 10%?

1.4 Refer to the oxygen concentration example in Sec. 1.7.2.

   (a) Given that nitrogen is lighter in weight than oxygen, is N_2 concentration at 10 km more or less that 25% of the sea level N_2 concentration?

   (b) What is the ratio of N_2 concentration to O_2 at 10 km? At sea level, the ratio is 4 to 1.

1.5 Show that the probability of an energy state being occupied ΔE above the Fermi level is the same as the probability of a state being empty ΔE below the Fermi level.

$$f(E_F + \Delta E) = 1 - f(E_F - \Delta E)$$

1.6 (a) Sketch the Fermi–Dirac distribution $f(E)$ at room temperature (300 K) and at a lower temperature such as 150 K. (Qualitative hand drawing.)

   (b) The state distribution in a system is given in Fig. 1–27, where each circle represents two electron states (one is spin-up; one is spin-down). Each electron state can be occupied by one electron. There is no state below E_{min}. The Fermi level at 0 K is given in Fig. 1–27. How many electrons are there in the system?
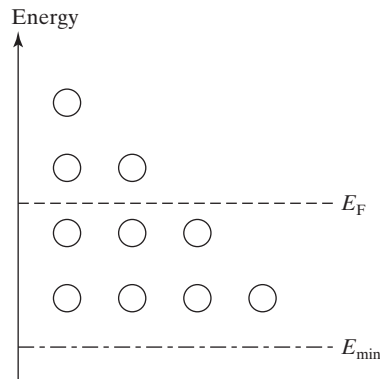

FIGURE 1–27

● **Energy: Density of States** ●

1.7 The carrier distributions in the conduction and valence bands were noted to peak at energies close to the band edges. (Refer to carrier distribution in Fig. 1–20.) Using Boltzmann approximation, show that the energy at which the carrier distribution peaks is $E_c + kT/2$ and $E_v - kT/2$ for the conduction and valence bands, respectively.

1.8 For a certain semiconductor, the densities of states in the conduction and valance bands are constants A and B, respectively. Assume non-degeneracy, i.e., E_F is not close to E_c or E_v.

   (a) Derive expressions for electron and hole concentrations.

   (b) If A = 2B, determine the location of the intrinsic Fermi energy (E_i) at 300 K with respect to the mid-bandgap of the semiconductor.
   Hint: These relationships may be useful:

$$\int_0^\infty x^{n-1} e^{-x} dx = \Gamma(n) \qquad \text{(Gamma function)}$$

$$\Gamma(2) = \Gamma(1) = 1, \Gamma(3) = 2, \Gamma(4) = 6$$

$$\Gamma(1/2) = \sqrt{\pi}, \ \Gamma(3/2) = 1/2\sqrt{\pi}, \ \Gamma(5/2) = 1/3\sqrt{\pi}.$$

1.9 For a certain semiconductor, the densities of states in the conduction and valence bands are: $D_c(E) = A \cdot (E - E_c) \cdot u(E - E_c)$ and $D_v(E) = B \cdot (E_v - E) \cdot u(E_v - E)$, respectively. $u(x)$, the unit step function, is defined as $u(x) = 0$ if $x < 0$ and $u(x) = 1$ if $x > 0$. Assume nondegeneracy, i.e. not too highly doped. You may find this fact useful:

$$\int_0^\infty x e^{-x} dx = 1$$

(a) Derive expressions for electron and hole concentrations as functions of the Fermi energy, E_F.

(b) If $A = 2B$, compute the intrinsic Fermi energy at 300 K.

1.10 The Maxwell–Boltzmann distribution function $f(E) = e^{-(E - E_f)/kT}$ is often used as an approximation to the Fermi–Dirac function. Use this approximation and the densities of the states in the conduction band $D_c(E) = A(E - E_c)^{1/2}$ to find:

(a) The energy at which one finds the most electrons ($1/cm^3 \cdot eV$).

(b) The conduction-band electron concentration (explain any approximation made).

(c) The ratio of the peak electron concentration at the energy of (a) to the electron concentration at $E = E_c + 40\,kT$ (about 1eV above E_c at 300 K). Does this result justify one of the approximations in part(b)?

(d) The average kinetic energy, $E - E_c$ of the electrons.
Hint: These relationships may be useful:

$$\int_0^\infty x^{n-1} e^{-x} dx = \Gamma(n) \qquad \text{(Gamma function)}$$

$$\Gamma(2) = \Gamma(1) = 1, \Gamma(3) = 2, \Gamma(4) = 6$$

$$\Gamma(1/2) = \sqrt{\pi},\ \Gamma(3/2) = 1/2\sqrt{\pi},\ \Gamma(5/2) = 3/4\sqrt{\pi}.$$

● **Electron and Hole Concentrations** ●

1.11 (a) The electron concentration in a piece of Si at 300 K is $10^5\,cm^{-3}$. What is the hole concentration?

(b) A semiconductor is doped with impurity concentrations N_d and N_a such that $N_d - N_a \gg n_i$ and all the impurities are ionized. Determine n and p.

(c) In a silicon sample at T = 300 K, the Fermi level is located at 0.26 eV (10 kT) above the intrinsic Fermi level. What are the hole and electron concentrations?

(d) What are the hole and electron concentration at T = 800 K for the sample in part (c), and where approximately is E_F? Comment on your results.

● **Nearly Intrinsic Semiconductor** ●

1.12 For a germanium sample at room temperature, it is known that $n_i = 10^{13}\,cm^{-3}$, $n = 2p$, and $N_a = 0$. Determine n and N_d.

1.13 Boron atoms are added to a Si film resulting in an impurity density of $4 \times 10^{16}\,cm^{-3}$.

(a) What is the conductivity type (N-type or P-type) of this film?

(b) What are the equilibrium electron and hole densities at 300 K and 600 K?

(c) Why does the mobile carrier concentration increase at high temperatures?

(d) Where is the Fermi level located if T = 600 K?

● **Incomplete Ionization of Dopants and Freeze-Out** ●

1.14 Suppose you have samples of Si, Ge, and Ge, and GaAs at T = 300 K, all with the same doping level of $N_d^+ - N_a^- = 3 \times 10^{15}$ /cm^3. Assuming all dopants are ionized, for which material is p most sensitive to temperature (the sensitivity of p is define by $\delta p/\delta T$)? What is your conclusion regarding the relation between E_g and temperature sensitivity of minority carrier concentration? Repeat the problem using $(\delta p/\delta T)/p$ as the definition of sensitivity.

1.15 An N-type sample of silicon has uniform density ($N_d = 10^{19}$/cm^{-3}) of arsenic, and a P-type silicon sample has a uniform density ($N_a = 10^{15}$ /cm^{-3}) of boron. For each sample, determine the following:

(a) The temperature at which the intrinsic concentration n_i exceeds the impurity density by factor of 10.

(b) The equilibrium minority-carrier concentrations at 300 K. Assume full ionization of impurities.

(c) The Fermi level relative to the valence–band edge E_v in each material at 300 K.

(d) The electron and hole concentrations and the Fermi level if both types of impurities are present in the same sample.

1.16 A silicon sample is doped with $N_d = 10^{17}$cm^{-3} of As atoms.

(a) What are the electron and hole concentrations and the Fermi level position (relative to E_c or E_v) at 300 K? (Assume full ionization of impurities.)

(b) Check the full ionization assumption using the calculated Fermi level, (i.e., find the probability of donor states being occupied by electrons and therefore not ionized.) Assume that the donor level lies 50 meV below the conduction band, (i.e., $E_c - E_D = 50$ meV.)

(c) Repeat (a) and (b) for $N_d = 10^{19}$cm^{-3}. (Discussion: when the doping concentration is high, donor (or acceptor) **band** is formed and that allows all dopant atoms to contribute to conduction such that "full ionization" is a good approximation after all).

(d) Repeat (a) and (b) for $N_d = 10^{17}$cm^{-3} but T = 30K. (This situation is called dopant freeze-out.)

1.17 Given N-type silicon sample with uniform donor doping of (a) $N_d = 10^{18}$/cm^3, (b) $N_d = 10^{19}$/cm^3, and (c) $N_d = 10^6$/cm^3, calculate the Fermi levels at room temperature assuming full ionization for all cases. Check whether the above assumption of full ionization of each case is correct with the calculated Fermi level. When this is not correct, what is the relative position of E_F and E_D? Assume that

$$E_c - E_D = 0.05 \text{ eV}.$$

● **REFERENCES** ●

1. Shockley, W. *Electrons and Holes in Semiconductors*. Princeton, NJ: Van Nostrand, 1950.

2. Shur, M. *Physics of Semiconductor Devices*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1990.

3. Neamen, D. *Semiconductor Physics and Devices,* 3rd ed. New York: McGraw-Hill, 2003.

4. Smith, R.A. *Semiconductors*, 2nd ed. London: Cambridge University Press, 1979.

5. Streetman, B.G. *Solid State Electronic Devices*, 6th ed. Upper Saddle River, NJ: Prentice-Hall, Inc., 2006.

6. Pierret, R. F. *Modular Series on Solid State Devices*, Vol. I. Reading, MA: Addison-Wesley Publishing Co., 1983.

7. Sze, S. M. *Physics of Semiconductor Devices*, 2nd ed. New York: John Wiley & Sons, 1981.

8. Taur, Y., and T. H. Ning. *Fundamentals of Modern VLSI Devices*. Cambridge, UK: Cambridge University Press, 1998.

● **GENERAL REFERENCES** ●

1. Neamen, D. *Semiconductor Physics and Devices,* 3rd ed. New York: McGraw-Hill, 2003.

2. Streetman, B.G. *Solid State Electronic Devices*, 6th ed. Upper Saddle River, NJ: Prentice-Hall, Inc., 2006.