IT IS WELL KNOWN THAT ARSENIC IS A POISON. IT IS LESS WELL KNOWN THAT SILICON CRYSTALS THAT HAVE SMALL CONCENTRATIONS OF ARSENIC ATOMS HAVE A MUCH LOWER RESISTIVITY THAN DO CRYSTALS THAT ARE 100 PERCENT SILICON. *(The Natural History Museum/Alamy.)*

? Do you know how many atoms of arsenic it takes to increase the charge-carrier density by a factor of 5 million? (See Example 38-7.)

Solids

The first microscopic model of electric conduction in metals was proposed by Paul K. Drude in 1900 and developed by Hendrik A. Lorentz about 1909. This model successfully predicts that the current is proportional to the potential drop (Ohm's law) and relates the resistivity of conductors to the mean speed and the mean free path* of the free electrons within the conductor. However, when mean speed and mean free path are interpreted classically, there is a disagreement between the calculated values and the measured values of the resistivity, and a similar disagreement between the predicted temperature dependence and the observed temperature dependence that resistivity values have. Thus, the classical theory fails to adequately describe the resistivity of metals. Furthermore, the classical theory says nothing about the most striking

---

* The mean free path is the average distance traveled between collisions.

1281

property of solids, namely, that some substances are conductors, others are insulators, and still others are semiconductors, which are substances whose resistivity falls between that of conductors and insulators.

When mean speed and mean free path are interpreted using quantum theory, both the magnitude and the temperature dependence of the resistivity are correctly predicted. In addition, quantum theory allows us to determine if a substance will be a conductor, an insulator, or a semiconductor.

In this chapter, we use our understanding of quantum mechanics to discuss the structure of solids and solid-state semiconducting devices. Much of our discussion will be qualitative because, as in atomic physics, the quantum-mechanical calculations are mathematically sophisticated.

38-1 THE STRUCTURE OF SOLIDS

The three phases of matter we observe everyday—gas, liquid, and solid—result from the relative strengths of the attractive forces between atoms and molecules and the thermal energies of the particles. Molecules and atoms in the gas phase have relatively large thermal kinetic energies, and such particles have little influence on one another except during their frequent but brief collisions. (By using the term thermal kinetic energies, we mean the kinetic energies of the molecules and atoms in the center-of-mass reference frame of the gas.) At sufficiently low temperatures, van der Waals forces will cause practically every substance to condense into a liquid and then into a solid. In liquids, the molecules or atoms are close enough—and their thermal kinetic energies are low enough—that they can develop a temporary **short-range order.** As their thermal kinetic energies are further reduced, the molecules or atoms form solids, which are characterized by a lasting order.

If a liquid is cooled slowly so that the kinetic energy of its molecules is reduced slowly, the molecules (or atoms or ions) may arrange themselves in a regular crystalline array, producing the maximum number of bonds and leading to a minimum potential energy. However, if the liquid is cooled rapidly so that its internal energy is removed before the molecules have a chance to arrange themselves, the solid formed is often not crystalline or the arrangement is not regular. Such a solid is called an **amorphous solid.** It displays short-range order but not the long-range order (the order over many molecular, atomic, or ionic diameters) that is characteristic of a crystal. Glass is a typical amorphous solid. A characteristic result of the long-range ordering of a crystal is that it has a well-defined melting point, whereas an amorphous solid merely softens as its temperature is increased. Many substances may solidify into either an amorphous state or a crystalline state depending on how the substances are prepared; others exist only in one such state or the other.

Most common solids are polycrystalline; that is, they consist of many single crystals that meet at *grain boundaries.* The size of a single crystal is typically a fraction of a millimeter. However, large single crystals do occur naturally and can be produced artificially. The most important property of a single crystal is the symmetry and regularity of its structure. It can be thought of as having a single unit structure that is repeated throughout the crystal. This smallest unit of a crystal is called the **unit cell;** its structure depends on the type of bonding—ionic, covalent, metallic, hydrogen, van der Waals—between the atoms, ions, or molecules. If more than one kind of atom is present, the structure will also depend on the relative sizes of the atoms.

Figure 38-1 shows the structure unit cell of crystalline sodium chloride (NaCl). The Na^+ and Cl^- ions are spherically symmetric, and the Cl^- ion is approximately twice as large as the Na^+ ion. The minimum potential energy for this crystal occurs when an ion of either kind has six nearest neighbors of the other kind. This structure is called *face-centered-cubic* (fcc). Note that the Na^+ and Cl^- ions in solid NaCl are *not* paired into NaCl molecules.

Cl^-

Na^+
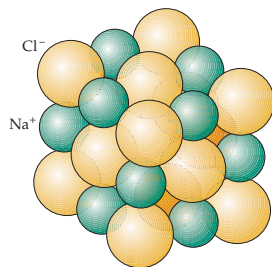
FIGURE 38-1 Face-centered-cubic structure of the NaCl crystal.

The net attractive part of the potential energy of an ion in a crystal can be written

$$U_{att} = -\alpha \frac{ke^2}{r}$$
38-1

where r is the (center-to-center) separation distance between neighboring ions (0.281 nm for the Na^+ and Cl^- ions in crystalline NaCl) and α, called the **Madelung constant,** depends on the geometry of the crystal. If only the six nearest neighbors of each ion in a face-centered-cubic crystalline structure were important, α would be six. However, in addition to the six neighbors of the opposite charge at a distance r, there are twelve ions of the same charge at a distance $\sqrt{2}r$, eight ions of opposite charge at a distance $\sqrt{3}r$, and so on. The Madelung constant is thus an infinite sum:

$$\alpha = 6 - \frac{12}{\sqrt{2}} + \frac{8}{\sqrt{3}} - \dots$$
38-2

The value of the Madelung constant for face-centered-cubic structures is $\alpha = 1.7476$.*

---

* A large number of terms are needed to calculate the Madelung constant accurately because the sum converges very slowly.
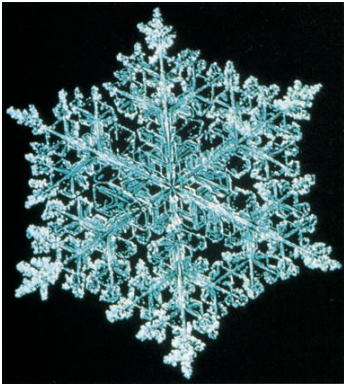
Crystal structure. (a) The hexagonal symmetry of a snowflake arises from a hexagonal symmetry in its lattice of hydrogen atoms and oxygen atoms. (b) NaCl (salt) crystals, magnified approximately thirty times. The crystals are built up from a cubic lattice of sodium and chloride ions. In the absence of impurities, an exact cubic crystal is formed. This (false-color) scanning electron micrograph shows that in practice the basic cube is often disrupted by dislocations, giving rise to crystals that have a wide variety of shapes. The underlying cubic symmetry, though, remains evident. (c) A crystal of quartz (SiO_2, silicon dioxide), the most abundant and widespread mineral on Earth. If molten quartz solidifies without crystallizing, glass is formed. (d) A soldering iron tip, ground down to reveal the copper core within its iron sheath. Visible in the iron is its underlying microcrystalline structure.
((a) Richard Waters 2/89 p. 52 Discover. (b) © Dr. Jeremy Burgess/Science Photo Library/Photo Researchers. (c) © Thomas R. Taylor/Photo Researchers. (d) Courtesy the AT&T Archives.)

(b)

(c)

(d)

When Na^+ and Cl^- ions are very close together, they repel each other because of the overlap of their electron orbitals and the exclusion-principle repulsion discussed in Section 37-1. A simple empirical expression for the potential energy associated with this repulsion that works fairly well is

$$U_{rep} = \frac{A}{r^n}$$

where A and n are constants. The total potential energy of an ion is then

$$U = -\alpha\frac{ke^2}{r} + \frac{A}{r^n} \qquad \text{38-3}$$

The equilibrium separation $r = r_0$ is that at which the force $F = -dU/dr$ is zero. Differentiating and setting $dU/dr = 0$ at $r = r_0$, we obtain

$$A = \frac{\alpha ke^2 r_0^{n-1}}{n} \qquad \text{38-4}$$

Substituting for A in Equation 38-3 gives

$$U = -\alpha\frac{ke^2}{r_0}\left[\frac{r_0}{r} - \frac{1}{n}\left(\frac{r_0}{r}\right)^n\right] \qquad \text{38-5}$$

At $r = r_0$, we have

$$U(r_0) = -\alpha\frac{ke^2}{r_0}\left(1 - \frac{1}{n}\right) \qquad \text{38-6}$$

If we know the equilibrium separation r_0, the value of n can be found approximately from the *dissociation energy* of the crystal, which is the energy needed to break up the crystal into atoms.

Example 38-1 **Separation Distance between Na^+ and Cl^- in NaCl**

Calculate the equilibrium separation r_0 for NaCl from the measured density of NaCl, which is $\rho = 2.16$ g/cm³.

PICTURE We consider each ion to occupy a cubic volume of side r_0. The mass of 1 mol of NaCl is 58.4 g, which is the sum of the molar masses of sodium and chlorine. There are $2N_A$ ions in 1 mol of NaCl, where $N_A = 6.02 \times 10^{23}$ is Avogadro's number.

SOLVE

1. We consider each ion to occupy a cubic volume of side r_0. The volume v of one mole of NaCl equals the number of ions multiplied by the volume per ion:

$$v = 2N_A r_0^3$$

2. Relate r_0 to the density ρ and the molar mass M of NaCl:

$$\rho = \frac{M}{v} = \frac{M}{2N_A r_0^3}$$

3. Solve for r_0^3 and substitute the known values:

$$r_0^3 = \frac{M}{2N_A\rho} = \frac{58.4 \text{ g}}{2(6.02 \times 10^{23})(2.16 \text{ g/cm}^3)}$$

$$= 2.25 \times 10^{-23} \text{ cm}^3$$

so

$$r_0 = 2.82 \times 10^{-8} \text{ cm} = \boxed{0.282 \text{ nm}}$$

CHECK In Chapter 36, we found the diameter of the hydrogen atom in the ground state to be about 0.11 nm. Our step 3 result is less than three times larger. Thus, $r_0 = 0.282$ nm is plausible.

The measured dissociation energy of NaCl is 770 kJ/mol. Using $1 \text{ eV} = 1.602 \times 10^{-19}$ J and the fact that 1 mol of NaCl has N_A pairs of ions, we can express the dissociation energy in electron volts per ion pair. The conversion between electron volts per ion pair and kilojoules per mole is

$$1\frac{\text{eV}}{\text{ion pair}} \times \frac{6.022 \times 10^{23} \text{ ion pairs}}{1 \text{ mol}} \times \frac{1.602 \times 10^{-19} \text{ J}}{1 \text{ eV}}$$

The result is

$$1 \frac{eV}{\text{ion pair}} = 96.47 \frac{kJ}{\text{mol}} \qquad \text{38-7}$$

Thus, 770 kJ/mol = 7.98 eV per ion pair. Substituting -7.98 eV for $U(r_0)$, 0.282 nm for r_0, and 1.75 for α in Equation 38-6, we can solve for n. The result is $n = 9.35 \approx 9$.

Most ionic crystals, such as LiF, KF, KCl, KI, and AgCl, have a face-centered-cubic structure. Some elemental solids that have fcc structure are silver, aluminum, gold, calcium, copper, nickel, and lead.

Figure 38-2 shows the structure of CsCl, which is called the *body-centered-cubic* (bcc) structure. In this structure, each ion has eight nearest neighbor ions of the opposite charge. The Madelung constant for these crystals is 1.7627. Elemental solids that have bcc structure include barium, cesium, iron, potassium, lithium, molybdenum, and sodium.
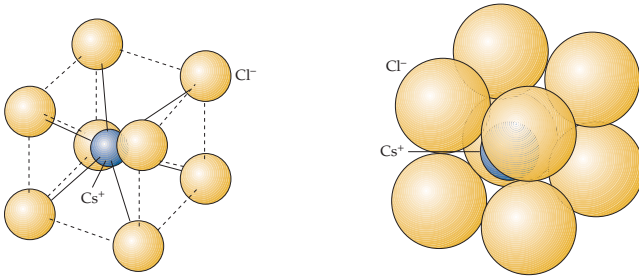
Figure 38-3 shows another important crystal structure: the *hexagonal close-packed* (hcp) structure. This structure is obtained by stacking identical spheres, such as bowling balls. In the first layer, each ball touches six others; thus, the name *hexagonal*. In the next layer, each ball fits into a triangular depression of the first layer. In the third layer, each ball fits into a triangular depression of the second layer, so it lies directly over a ball in the first layer. Elemental solids that have hcp structure include beryllium, cadmium, cerium, magnesium, osmium, and zinc.

For solids that have covalent bonding, the crystal structure is determined by the configuration of the bonds. Figure 38-4 illustrates the diamond structure of carbon, in which each atom is bonded to four other atoms as a result of hybridization, which is discussed in Section 37-2. This configuration is also the structure of germanium and silicon.
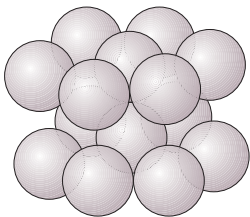

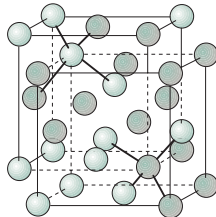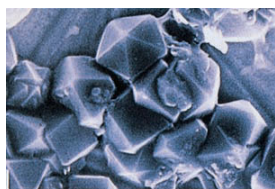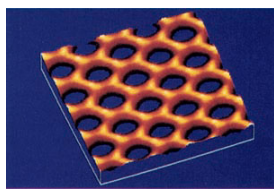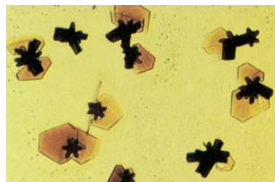**F I G U R E   38-3** Hexagonal close-packed crystal structure.

F I G U R E   38-4 Diamond crystal structure. This structure can be considered to be a combination of two interpenetrating face-centered-cubic structures.
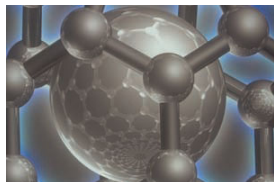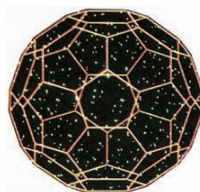
(a)

(b)

(c)

(d)

(e)

Carbon exists in three well-defined crystalline forms: diamond, graphite, and fullerenes (short for "buckminsterfullerenes"). Fullerenes were discovered in 1985. The forms differ in how the carbon atoms are packed together in a lattice. A fourth form of carbon, in which no well-defined crystalline form exists, is common charcoal. (a) Synthetic diamonds, magnified approximately 75,000 times. In diamond, each carbon atom is centered in a tetrahedron of four other carbon atoms. The strength of these bonds accounts for the hardness of a diamond. (b) An atomic-force micrograph of graphite. In graphite, carbon atoms are arranged in sheets, where each sheet is made up of atoms in hexagonal rings. The sheets slide easily across one another, a property that allows graphite to function as a lubricant. (c) A single sheet of carbon rings can be closed on itself if certain rings are allowed to be pentagonal, instead of hexagonal. A computer-generated image of the smallest such structure, C_{60}, is shown here. Each of the sixty vertices corresponds to a carbon atom; twenty of the faces are hexagons and twelve of the faces are pentagons. The same geometric pattern is encountered in a soccer ball. (d) Fullerene crystals, in which C_{60}

molecules are close-packed. The smaller crystals tend to form thin brownish platelets; larger crystals are usually rodlike in shape. Fullerenes exist in which more than sixty carbon atoms appear. In the crystals shown here, about one-sixth of the molecules are C_{70}. (e) Carbon nanotubes have very interesting electrical properties. A single graphite sheet is a semimetal, which means that it has properties intermediate between those of semiconductors and those of metals. When a graphite sheet is rolled into a nanotube, not only do the carbon atoms have to line up around the circumference of the tube, but the wave functions of the electrons must also match up. This boundary-matching requirement places restrictions on these wave functions, which affects the motion of the electrons. Depending on exactly how the tube is rolled up, the nanotube can be either a semiconductor or a metal. *((a) Chris Kovach 3/91 p. 69 Discover. (b) Srinivas Manne, University of California, Santa Barbara. (c) Dr. F. A. Quiocho and J. S. Spurlino/Howard Hughes Medical Institute, Baylor College of Medicine. (d) W. Krätschmer/Max-Planck-Institute for Nuclear Physics. (e) © Kenneth Weard/BioGrafx/Science Source/Photo Researchers.)*

38-2 A MICROSCOPIC PICTURE OF CONDUCTION

We consider a metal as a regular three-dimensional lattice of ions filling some volume V and having a large number N of electrons that are free to move throughout the whole metal. The number of free electrons in a metal is approximately one to four electrons per atom. In the absence of an electric field, the free electrons move about the metal randomly, much the way gas molecules move about in a container.

The current in a conducting wire segment is proportional to the voltage drop across the segment:

$$I = \frac{V}{R} \qquad \text{(or } V = IR\text{)}$$

The resistance R is proportional to the length L of the wire segment and inversely proportional to the cross-sectional area A:

$$R = \rho \frac{L}{A}$$

where ρ is the resistivity. Substituting $\rho L/A$ for R, and EL for V, we can write the current in terms of the electric field strength E and the resistivity. We have

$$I = \frac{V}{R} = \frac{EL}{\rho L/A} = \frac{1}{\rho}EA$$

Dividing both sides by the area A gives $I/A = (1/\rho)E$, or $J = (1/\rho)E$, where $J = I/A$ is the magnitude of the **current density** vector $\vec{J}$. The current density vector is defined as

$$\vec{J} = qn\vec{v}_d \qquad \text{38-8}$$

DEFINITION—CURRENT DENSITY

where q, n, and $\vec{v}_d$ are the charge, the number density, and the drift velocity of the charge carrier. (This follows from Equation 25-3.) In vector form, the relation between the current density and the electric field is

$$\vec{J} = \frac{1}{\rho}\vec{E} \qquad \text{38-9}$$

This relation is the point form of Ohm's law. The reciprocal of the resistivity is called the **conductivity.**

According to Ohm's law, the resistivity is independent of both the current density and the electric field $\vec{E}$. Combining Equations 38-8 and 38-9 gives

$$-en_e\vec{v}_d = \frac{1}{\rho}\vec{E} \qquad \text{38-10}$$

where $-e$ and n_e have been substituted for q and n, respectively. According to Equation 38-10, the drift velocity $\vec{v}_d$ is proportional to $\vec{E}$.

In the presence of an electric field, a free electron experiences a force $-e\vec{E}$. If this were the only force acting, the electron would have a constant acceleration $-e\vec{E}/m_e$. However, Equation 38-10 implies a steady-state situation with a constant drift velocity that is proportional to the field $\vec{E}$. In the microscopic model, it is assumed that a free electron is accelerated for a short time and then makes a collision with a lattice ion. The velocity of the electron immediately after the collision is completely unrelated to the drift velocity. The justification for this assumption is that the magnitude of the drift velocity is extremely small compared with the speeds associated with the thermal kinetic energies of the free electrons.

For a typical free electron, its velocity a time t after its last collision is $\vec{v}_0 - (-e\vec{E}/m_e)t$, where $\vec{v}_0$ is its velocity immediately after that collision. Because the direction of $\vec{v}_0$ is random, it does not contribute to the average velocity of the electrons. Thus, the average velocity or drift velocity of the electrons is

$$\vec{v}_d = -\frac{e\vec{E}}{m_e}\tau \qquad \text{38-11}$$

where τ is the average time since the last collision. Substituting for $\vec{v}_d$ in Equation 38-10, we obtain

$$-n_e e\left(\frac{e\vec{E}}{m_e}\tau\right) = \frac{1}{\rho}\vec{E}$$

so

$$\rho = \frac{m_e}{n_e e^2 \tau} \qquad \text{38-12}$$

The time τ, called the **collision time,** is also the average time between collisions.*

* It is tempting but incorrect to think that if τ is the average time between collisions, the average time since its last collision is $\frac{1}{2}\tau$ rather than τ. If you find this confusing, you may take comfort in the fact that Drude used the incorrect result $\frac{1}{2}\tau$ in his original work.

The average distance an electron travels between collisions is $v_{av}\tau$, which is called the mean free path λ:

$$\lambda = v_{av}\tau \qquad \text{38-13}$$

where v_{av} is the mean speed of the electrons. (The mean speed is many orders of magnitude greater than the drift speed.) In terms of the mean free path and the mean speed, the resistivity is

$$\rho = \frac{m_e v_{av}}{n_e e^2 \lambda} \qquad \text{38-14}$$

RESISTIVITY IN TERMS OF v_{AV} AND λ

According to Ohm's law, the resistivity ρ is independent of the electric field $\vec{E}$. Because m_e, n_e, and e are constants, the only quantities that could possibly depend on $\vec{E}$ are the mean speed v_{av} and the mean free path λ. Let us examine these quantities to see if they can possibly depend on the applied field $\vec{E}$.

CLASSICAL INTERPRETATION OF v_{av} AND λ

Classically, at $T = 0$ all the free electrons in a conductor should have zero kinetic energy. As the conductor is heated, the lattice ions acquire an average kinetic energy of $\frac{3}{2}kT$, which is imparted to the free electrons by the collisions between the electrons and the ions. (This is a result of the equipartition theorem studied in Chapters 17 and 18.) The free electrons would then have a Maxwell–Boltzmann distribution just like a gas of molecules. In equilibrium, the electrons would be expected to have a mean kinetic energy of $\frac{3}{2}kT$, which at ordinary temperatures (~300 K) is approximately 0.04 eV. At $T = 300$ K, their root-mean-square (rms) speed,* which is slightly greater than the mean speed, is

$$v_{av} \approx v_{rms} = \sqrt{\frac{3kT}{m_e}} = \sqrt{\frac{3(1.38 \times 10^{-23}\,\text{J/K})(300\,\text{K})}{9.11 \times 10^{-31}\text{kg}}}$$
$$= 1.17 \times 10^5\,\text{m/s} \qquad \text{38-15}$$

Note that this is about nine orders of magnitude greater than the typical drift speed of 3.5×10^{-5} m/s, which was calculated in Example 25-1. The very small drift speed caused by the electric field therefore has essentially no effect on the very large mean speed of the electrons, so v_{av} in Equation 38-14 cannot depend on the electric field $\vec{E}$.

The mean free path is related classically to the size of the lattice ions in the conductor and to the number of ions per unit volume. Consider one electron moving with speed v through a region of stationary ions that are assumed to be hard spheres (Figure 38-5). Assume the size of the electron is negligible. The electron will collide with an ion if it comes within a distance r from the center of the ion, where r is the radius of the ion. During some time interval Δt_1, the electron moves a distance vt_1. If there is an ion whose center is in the cylindrical volume $\pi r^2 v\Delta t_1$, the electron will collide with the ion. The electron will then change directions and collide with another ion in time Δt_2 if the center of the ion is in the volume $\pi r^2 vt_2$. Thus, in the total time $\Delta t = \Delta t_1 + \Delta t_2 + \dots$, the electron will collide with the number of ions whose centers are in the volume $\pi r^2 v\Delta t$. The number of ions in this volume is $n_{ion}\pi r^2 v\Delta t$, where n_{ion} is the number of ions per unit volume.
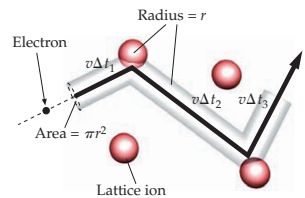
FIGURE 38-5 Model of an electron moving through the lattice ions of a conductor. The electron, which is considered to be a point particle, collides with an ion if it comes within a distance r of the center of the ion, where r is the radius of the ion. If the electron speed is v, it collides in time Δt with all the ions whose centers are in the volume $\pi r^2 v\Delta t$. While this picture is in accord with the classical Drude model for conduction in metals, it is in conflict with the current quantum-mechanical model presented later in this chapter.

---

* See Equation 17–21.

The total path length divided by the number of collisions is the mean free path:

$$\lambda = \frac{v\Delta t}{n_{\text{ion}}\pi r^2 v\Delta t} = \frac{1}{n_{\text{ion}}\pi r^2} = \frac{1}{n_{\text{ion}}A} \qquad \text{38-16}$$

where $A = \pi r^2$ is the cross-sectional area of a lattice ion.

SUCCESSES AND FAILURES OF THE CLASSICAL MODEL

Neither n_{ion} nor r depends on the electric field $\vec{E}$, so λ also does not depend on $\vec{E}$. v_{av} and λ do not depend on $\vec{E}$ according to their classical interpretations, so the resistivity ρ does not depend on $\vec{E}$ in accordance with Ohm's law. However, the classical theory gives an incorrect temperature dependence for the resistivity. Because λ depends only on the radius and the number density of the lattice ions, the only quantity in Equation 38-14 that depends on temperature in the classical theory is v_{av}, which is proportional to $\sqrt{T}$. But experiments show that ρ varies linearly with temperature. Furthermore, when ρ is calculated at $T = 300$ K using the Maxwell–Boltzmann distribution for v_{av} and Equation 38-16 for λ, the calculated result is about six times greater than the measured value.

The classical theory of conduction fails because electrons are not classical particles. The wave nature of the electrons must be considered. Because of the wave properties of electrons and the constraints described by the exclusion principle (to be discussed in the following section), the energy distribution of the free electrons in a metal is not even approximately given by the Maxwell–Boltzmann distribution. Furthermore, the collision of an electron with a lattice ion is not similar to the collision of a baseball with a tree. Instead, it involves the scattering of electron waves by the lattice. To understand the quantum theory of conduction, we need a qualitative understanding of the energy distribution of free electrons in a metal. This will also help us understand the origin of contact potentials between two dissimilar metals in contact and the contribution of free electrons to the heat capacity of metals.

38-3 FREE ELECTRONS IN A SOLID

One may want to consider free electrons in a metal to be an *electron gas* in a metal. However, molecules in an ordinary gas, such as air, obey the classical Maxwell–Boltzmann energy distribution, but the free electrons in a metal do not. Instead, they obey a quantum energy distribution called the *Fermi–Dirac distribution.* The main features of a free electron can be understood by considering the electron in a metal to be a particle in a box, a problem whose one-dimensional version we studied extensively in Chapter 34. We discuss the main features of a free electron semiquantitatively in this section and leave the details of the Fermi–Dirac distribution to Section 38-9.

ENERGY QUANTIZATION IN A BOX

In Chapter 34, we found that the wavelength associated with an electron of momentum p is given by the de Broglie relation:

$$\lambda = \frac{h}{p} \qquad \text{38-17}$$

where h is Planck's constant. When a particle is confined to a finite region of space, such as a box, only certain wavelengths λ_n, where $n = 1, 2, \ldots$, that are specified by standing-wave conditions are allowed. For a one-dimensional box of length L, the standing-wave condition is

$$n\frac{\lambda_n}{2} = L \qquad n = 1, 2, \ldots \qquad \text{38-18}$$

This results in the quantization of energy:

$$E_n = \frac{p_n^2}{2m} = \frac{(h/\lambda_n)^2}{2m} = \frac{h^2}{2m}\frac{1}{\lambda_n^2} = \frac{h^2}{2m}\frac{1}{(2L/n)^2}$$

or

$$E_n = n^2 E_1 \qquad\qquad 38\text{-}19$$

where $E_1 = h^2/(8mL^2)$. The spatial wave function for the nth state is given by

$$\psi_n(x) = \sqrt{\frac{2}{L}}\sin(n\pi x/L) \qquad\qquad 38\text{-}20$$

The quantum number n characterizes the wave function for a particular state and the energy of that state. In three-dimensional problems, three quantum numbers arise, one associated with each dimension.

THE EXCLUSION PRINCIPLE

The distribution of electrons among the possible energy states is described by the exclusion principle, which states that no two electrons in an atom can be in the same quantum state; that is, they cannot have the same set of values for their quantum numbers. The exclusion principle applies to all "spin one-half" particles (fermions), which include electrons, protons, and neutrons. These particles have a *spin* quantum number m_s which has two possible values, $+\frac{1}{2}$ and $-\frac{1}{2}$. The quantum state of a particle is characterized by the spin quantum number m_s and the quantum numbers associated with the spatial part of the wave function. Because the spin quantum numbers have just two possible values, the exclusion principle can be stated in terms of the spatial states:

> There can be at most two electrons with the same set of values for their *spatial* quantum numbers.
>
> EXCLUSION PRINCIPLE IN TERMS OF SPATIAL STATES

When there are more than two electrons in a system, such as an atom, only two can be in the lowest energy state. The third and fourth electrons must go into the second-lowest state, and so on.

Example 38-2 | **Boson-System Energy versus Fermion-System Energy**

Compare the total energy of the ground state of five identical bosons of mass m in a one-dimensional box with the total energy of the ground state of five identical fermions of mass m in the same box.

PICTURE The ground state is the lowest possible energy state. The energy levels in a one-dimensional box are given by $E_n = n^2 E_1$, where $E_1 = h^2/(8mL^2)$. (This is in accord with Equation 38-19.) The lowest energy for five bosons occurs when all the bosons are in the state $n = 1$, as shown in Figure 38-6a. For fermions, the lowest state occurs when two fermions are in the state $n = 1$, two fermions are in the state $n = 2$, and one fermion is in the state $n = 3$, as shown in Figure 38-6b.
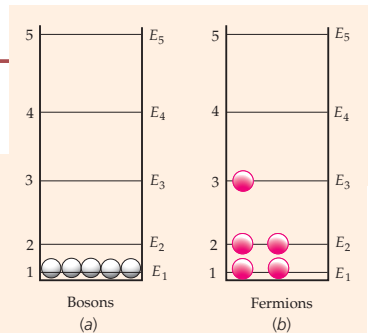
FIGURE 38-6

Bosons
(a)

Fermions
(b)

SOLVE

1. The energy of five bosons in the state $n = 1$ is:

$E = 5E_1$

2. The energy of two fermions in the state $n = 1$, two fermions in the state $n = 2$, and one fermion in the state $n = 3$ is:

$E = 2E_1 + 2E_2 + 1E_3 = 2E_1 + 2(2)^2E_1 + 1(3)^2E_1$
$= 2E_1 + 8E_1 + 9E_1 = 19E_1$

3. Compare the total energies:

The five identical fermions have 3.8 times the total energy of the five identical bosons.

CHECK The fact that fermions must have different quantum states has a large effect on the total energy of a multiple-particle system, as expected.

THE FERMI ENERGY

When there are many electrons in a box, at $T = 0$ the electrons will occupy the lowest energy states consistent with the exclusion principle. If we have N electrons, we can put two electrons in the lowest energy level, two electrons in the next lowest energy level, and so on. The N electrons thus fill the lowest $N/2$ energy levels (Figure 38-7). The energy of the last filled (or half-filled) level at $T = 0$ is called the Fermi energy E_F. If the electrons moved in a one-dimensional box, the Fermi energy would be given by Equation 38-19, with $n = N/2$:

$$E_F = \left(\frac{N}{2}\right)^2 \frac{h^2}{8m_e L^2} = \frac{h^2}{32m_e}\left(\frac{N}{L}\right)^2 \qquad 38\text{-}21$$

FERMI ENERGY AT $T = 0$ IN ONE DIMENSION

In a one-dimensional box, the Fermi energy depends on the number of free electrons per unit length of the box.

FIGURE 38-7 At $T = 0$ the electrons fill up the allowed energy states to the Fermi energy E_F. The levels are so closely spaced that they can be assumed to be continuous.

PRACTICE PROBLEM 38-1

Suppose there is an ion, and therefore a free electron, every 0.100 nm in a one-dimensional box. Calculate the Fermi energy. *Hint: Write Equation 38-21 as*

$$E_F = \frac{(hc)^2}{32m_e c^2}\left(\frac{N}{L}\right)^2 = \frac{(1240\text{ eV}\cdot\text{nm})^2}{32(0.511\text{ MeV})}\left(\frac{N}{L}\right)^2$$

In our model of conduction, the free electrons move in a *three-dimensional* box of volume V. The derivation of the Fermi energy in three dimensions is somewhat difficult, so we will just give the result. In three dimensions, the Fermi energy at $T = 0$ is
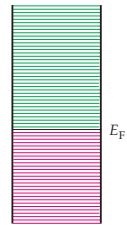
$$E_F = \frac{h^2}{8m_e}\left(\frac{3N}{\pi V}\right)^{2/3} \qquad 38\text{-}22a$$

FERMI ENERGY AT $T = 0$ IN THREE DIMENSIONS

The Fermi energy depends on the number density of free electrons N/V. Substituting numerical values for the constants gives

$$E_F = (0.3646\text{ eV}\cdot\text{nm}^2)\left(\frac{N}{V}\right)^{2/3} \qquad 38\text{-}22b$$

FERMI ENERGY AT $T = 0$ IN THREE DIMENSIONS

| Example 38-3 | **The Fermi Energy for Copper** |

The number density for electrons in copper was calculated in Example 25-1 and found to be 84.7/nm³. Calculate the Fermi energy at $T = 0$ for copper.

PICTURE The Fermi energy is given by Equations 38-22.

SOLVE

1. The Fermi energy is given by Equation 38-22b:

$$E_F = (0.3646 \text{ eV} \cdot \text{nm}^2)\left(\frac{N}{V}\right)^{2/3}$$

2. Substitute the given number density for copper:

$$E_F = (0.3646 \text{ eV} \cdot \text{nm}^2)(84.7/\text{nm}^3)^{2/3}$$
$$= \boxed{7.03 \text{ eV}}$$

CHECK The Fermi energy (the step-2 result) is much greater than kT at room temperatures as expected. For example, at $T = 300$ K, kT is only about 0.026 eV.

PRACTICE PROBLEM 38-2 Use Equation 38-22b to calculate the Fermi energy at $T = 0$ for gold, which has a free-electron number density of 59.0/nm³.

Table 38-1 lists the free-electron number densities and Fermi energies at $T = 0$ for several metals.

The free electrons in a metal are sometimes referred to as a Fermi gas. (They constitute a gas of fermions.) The average energy of a free electron can be calculated from the complete energy distribution of the electrons, which is discussed in Section 38-9. At $T = 0$, the average energy turns out to be

$$E_{av} = \tfrac{3}{5}E_F \qquad\qquad 38\text{-}23$$

AVERAGE ENERGY OF ELECTRONS IN A FERMI GAS AT $T = 0$

Table 38-1	**Free-Electron Number Densities* and Fermi Energies at $T = 0$ for Selected Elements**		
	Element	N/V, electrons/nm³	E_F, eV
Al	Aluminum	181	11.7
Ag	Silver	58.6	5.50
Au	Gold	59.0	5.53
Cu	Copper	84.7	7.03
Fe	Iron	170	11.2
K	Potassium	14.0	2.11
Li	Lithium	47.0	4.75
Mg	Magnesium	86.0	7.11
Mn	Manganese	165	11.0
Na	Sodium	26.5	3.24
Sn	Tin	148	10.2
Zn	Zinc	132	9.46

* Number densities are measured using the Hall effect, discussed in Section 26-4.

For copper, E_{av} is approximately 4 eV. This average energy is huge compared with thermal energies of about $kT \approx 0.026$ eV at a temperature of $T = 300$ K. This result is very different from the classical Maxwell–Boltzmann distribution result that at $T = 0, E = 0$, and that at some temperature T, E is of the same order as kT.

THE FERMI FACTOR AT $T = 0$

The probability of an energy state being occupied is called the **Fermi factor,** $f(E)$. At $T = 0$ all the states below E_F are filled, whereas all those above that energy are empty, as shown in Figure 38-8. Thus, at $T = 0$ the Fermi factor is simply

$$f(E) = \begin{cases} 1 & E < E_F \\ 0 & E > E_F \end{cases} \qquad 38\text{-}24$$
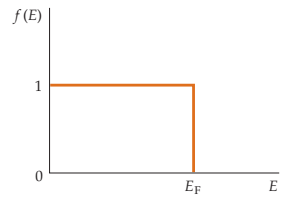
FIGURE 38-8 Fermi factor versus energy at $T = 0$.

THE FERMI FACTOR FOR $T > 0$

At temperatures greater than $T = 0$, some electrons will occupy higher energy states because of thermal energy gained during collisions with the lattice. However, an electron cannot move to a higher or lower state unless it is unoccupied. Because the kinetic energy of the lattice ions is of the order of kT, electrons cannot gain much more energy than kT in collisions with the lattice ions. Therefore, only those electrons that have energies within about kT of the Fermi energy can gain energy as the temperature is increased. At 300 K, kT is only 0.026 eV, so the exclusion principle prevents all but a very few electrons near the top of the energy distribution from gaining energy through random collisions with the lattice ions. Figure 38-9 shows a plot of the Fermi factor for some temperature T. Because for $T > 0$ there is no distinct energy that separates filled levels from unfilled levels, the definition of the Fermi energy must be slightly modified. At temperature T, the Fermi energy is defined to be the energy of the energy state for which the probability of being occupied is $\frac{1}{2}$. For all but extremely high temperatures, the difference between the Fermi energy at temperature T and the Fermi energy at temperature $T = 0$ is very small.
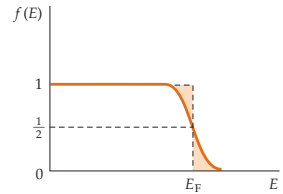
The **Fermi temperature** T_F is defined by

$$kT_F = E_F \qquad 38\text{-}25$$

For temperatures much lower than the Fermi temperature, the average energy of the lattice ions will be much less than the Fermi energy, and the electron energy distribution will not differ greatly from that at $T = 0$.

FIGURE 38-9 The Fermi factor for some temperature T. Some electrons that have energies near the Fermi energy are excited, as indicated by the shaded regions. The Fermi energy E_F is that value of E for which $f(E) = \frac{1}{2}$.

Example 38-4 — The Fermi Temperature for Copper

Find the Fermi temperature for copper.

PICTURE We use Equation 38-25 to find the Fermi temperature. The Fermi energy for copper at $T = 0$, calculated in Example 38-3, is 7.03 eV.

SOLVE

Use $E_F = 7.03$ eV and $k = 8.617 \times 10^{-5}$ eV/K in Equation 38-25:

$$T_F = \frac{E_F}{k} = \frac{7.03 \text{ eV}}{8.617 \times 10^{-5} \text{ eV/K}} = \boxed{81\,600 \text{ K}}$$

CHECK The Fermi temperature is very high, as expected.

TAKING IT FURTHER We can see from this example that the Fermi temperature of copper is much greater than any temperature T for which copper remains a solid.

Because an electric field in a conductor accelerates all of the conduction electrons together, the exclusion principle does not prevent the free electrons in filled states from participating in conduction. Figure 38-10 shows the Fermi factor in one dimension versus *velocity* for an ordinary temperature. The factor is approximately 1 for velocities v_x in the range $-u_F < v_x < u_F$, where the Fermi speed u_F is related to the Fermi energy by $E_F = \frac{1}{2}mu_F^2$. Then

$$u_F = \sqrt{\frac{2E_F}{m_e}}$$ 38-26
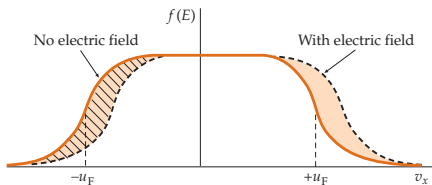
FIGURE 38-10 Fermi factor versus velocity in one dimension with no electric field (solid) and with an electric field in the $-x$ direction (dashed). The difference is greatly exaggerated.

Example 38-5 | **The Fermi Speed for Copper**

Calculate the Fermi speed for copper.

PICTURE We use Equation 38-26 to find the Fermi speed. The Fermi energy for copper at $T = 0$, calculated in Example 38-3, is 7.03 eV.

SOLVE

Use Equation 38-26 with $E_F = 7.03$ eV:

$$u_F = \sqrt{\frac{2(7.03\text{ eV})}{9.11 \times 10^{-31}\text{ kg}}\left(\frac{1.60 \times 10^{-19}\text{ J}}{1\text{ eV}}\right)} = \boxed{1.57 \times 10^6\text{ m/s.}}$$

**CHECK** As expected, the result (the Fermi speed for copper) is high, but less than the speed of light.

The dashed curve in Figure 38-10 shows the Fermi factor after the electric field has been acting for some time t. Although all of the free electrons have their velocities shifted in the direction opposite to the electric field, the net effect is equivalent to shifting only the electrons near the Fermi energy.

CONTACT POTENTIAL

When two different metals are placed in contact, a potential difference $V_{contact}$ called the **contact potential** develops between them. The contact potential depends on both the work functions of the two metals, ϕ_1 and ϕ_2 (we encountered work functions when the photoelectric effect was introduced in Chapter 34), and the Fermi energies of the two metals. When the metals are in contact, the total energy of the system is lowered if electrons near the boundary move from the metal that has the higher Fermi energy into the metal that has the lower Fermi energy until the Fermi energies of the two metals are the same, as shown in Figure 38-11. When equilibrium is established, the metal that has the lower initial Fermi energy is negatively charged and the other metal is positively charged, so that between them there is a potential difference $V_{contact}$ given by

$$V_{contact} = \frac{\phi_1 - \phi_2}{e}$$ 38-27

Table 38-2 lists the work functions for several metals.
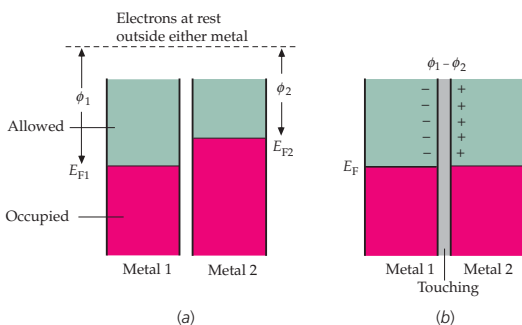
FIGURE 38-11 (a) Energy levels for two different metals that have different Fermi energies E_F and work functions ϕ. The work function is the difference between the energy of an electron at rest outside the metal and the Fermi energy within the metal. (b) When the metals are in contact, electrons flow from the metal that initially has the higher Fermi energy to the metal that initially has the lower Fermi energy until the Fermi energies are equal.

Table 38-2   Work Functions for Some Metals

	Metal	ϕ, eV		Metal	ϕ, eV
Ag	Silver	4.7	K	Potassium	2.1
Au	Gold	4.8	Mn	Manganese	3.8
Ca	Calcium	3.2	Na	Sodium	2.3
Cu	Copper	4.1	Ni	Nickel	5.2

Example 38-6   Contact Potential between Silver and Tungsten

The threshold wavelength for the photoelectric effect is 271 nm for tungsten and 262 nm for silver. What is the contact potential developed when silver and tungsten are placed in contact?

PICTURE The contact potential is proportional to the difference in the work functions for the two metals (Equation 38-27). The work function ϕ can be found from the given threshold wavelengths using $\phi = hc/\lambda_t$ (Equation 34-4).

SOLVE

1. The contact potential is given by Equation 38-27:

$$V_{contact} = \frac{\phi_1 - \phi_2}{e}$$

2. The work function is related to the threshold wavelength (Equation 34-4):

$$\phi = \frac{hc}{\lambda_t}$$

3. Substitute $\lambda_t = 271$ nm for tungsten (the symbol for tungsten is W):

$$\phi_W = \frac{hc}{\lambda_t} = \frac{1240 \text{ eV} \cdot \text{nm}}{271 \text{ nm}} = 4.58 \text{ eV}$$

4. Substitute $\lambda_t = 262$ nm for silver:

$$\phi_{Ag} = \frac{1240 \text{ eV} \cdot \text{nm}}{262 \text{ nm}} = 4.73 \text{ eV}$$

5. The contact potential is thus:

$$V_{contact} = \frac{\phi_{Ag} - \phi_W}{e} = 4.73 \text{ V} - 4.58 \text{ V} = \boxed{0.15 \text{ V}}$$

CHECK As expected, the contact potential is small (less than one volt). You do not get large potential differences just by putting two metals in contact.

HEAT CAPACITY DUE TO ELECTRONS IN A METAL

The quantum-mechanical description of the electron distribution in metals allows us to understand why the contribution of the free electrons to the heat capacity of a metal is much less that of the ions. According to the classical equipartition theorem, the energy of the lattice ions in n moles of a solid is $3nRT$, and thus the molar specific heat is $c' = 3R$, where R is the universal gas constant (see Section 18-7). In a metal, the number of free electrons is approximately equal to the number of lattice ions. If these electrons obey the classical equipartition theorem, they should have an energy of $\frac{3}{2}nRT$ and contribute an additional $\frac{3}{2}R$ to the molar specific heat. But measured heat capacities of metals are just slightly greater than those of insulators. We can understand this result because at some temperature T, only those electrons that have energies near the Fermi energy can be excited by random collisions with the lattice ions. The number of those electrons is of the order of $(kT/E_F)N$, where N is the total number of free electrons. The energy of those electrons is increased from that at $T = 0$ by an amount that is of the order of kT. So the total increase in thermal energy is of the order of $(kT/E_F)N \times kT$.

We can thus express the energy of N electrons at temperature T as

$$E = NE_{av}(0) + \alpha N \frac{kT}{E_F} kT \qquad 38\text{-}28$$

where $E_{av}(0)$ is the average energy at $T = 0$ and α is a constant that we expect to be of the order of 1 if our reasoning is correct. The calculation of α is quite challenging. The result is $\alpha = \pi^2/4$. Using this result and writing E_F in terms of the Fermi temperature, $E_F = kT_F$, we obtain the following for the contribution of the free electrons to the heat capacity at constant volume:

$$C_V = \frac{dE}{dT} = 2\alpha Nk\frac{kT}{E_F} = \frac{1}{2}\pi^2 nR\frac{T}{T_F}$$

where we have written Nk in terms of the gas constant R $(R = Nk/n)$. The molar specific heat at constant volume is then

$$c'_V = \frac{1}{2}\pi^2 R\frac{T}{T_F} \qquad 38\text{-}29$$

We can see that because of the large value of T_F, the contribution of the free electrons is a small fraction of R at ordinary temperatures. Because $T_F = 81\,600$ K for copper, the molar specific heat of the free electrons at $T = 300$ K is

$$c'_V = \frac{1}{2}\pi^2\frac{300\text{ K}}{81\,600\text{ K}}R \approx 0.02R$$

which is in good agreement with the experiment.

38-4 QUANTUM THEORY OF ELECTRICAL CONDUCTION

We can use Equation 38-14 for the resistivity if we use the Fermi speed u_F (Equation 38-26) in place of v_{av}:

$$\rho = \frac{m_e u_F}{n_e e^2 \lambda} \qquad 38\text{-}30$$

We now have two problems. First, because the Fermi speed u_F is approximately independent of temperature, the resistivity given by Equation 38-30 is independent of temperature unless the mean free path should depend on the temperature. The second problem concerns magnitudes. As mentioned earlier, the classical expression for resistivity using v_{av} calculated from the Maxwell–Boltzmann distribution gives values that are about 6 times too large at $T = 300$ K. Because the Fermi speed u_F is about 16 times the Maxwell-Boltzmann value of v_{av}, the magnitude of ρ predicted by Equation 38-30 will be approximately 100 times greater than the experimentally determined value. The resolution of both of these problems lies in the calculation of the mean free path λ.

THE SCATTERING OF ELECTRON WAVES

In Equation 38-16 for the classical mean free path $\lambda = 1/(n_{ion}A)$, the quantity $A = \pi r^2$ is the cross-sectional area of the lattice ion as seen by an electron. In the quantum calculation, the mean free path is related to the scattering of electron waves by the crystal lattice. Detailed calculations show that, for a *perfectly* ordered crystal, $\lambda = \infty$; that is, there is no scattering of the electron waves. The scattering of electron waves arises because of *imperfections* in the crystal lattice, which have nothing to do with the actual cross-sectional area A of the lattice ions.

According to the quantum theory of electron scattering, A depends merely on *deviations* of the lattice ions from a perfectly ordered array and not on the size of the ions. The most common causes of such deviations are thermal vibrations of the lattice ions or impurities.

We can use $\lambda = 1/(n_{\text{ion}}A)$ for the mean free path if we reinterpret the area A. Figure 38-12 compares the classical picture and the quantum picture of this area. In the quantum picture, the lattice ions are points that have no size but present an area $A = \pi r_0^2$, where r_0 is the amplitude of thermal vibrations. In Chapter 14, we saw that the energy of vibration in simple harmonic motion is proportional to the square of the amplitude, which is πr_0^2. Thus, the effective area A is proportional to the energy of vibration of the lattice ions. From the equipartition theorem,* we know that the average energy of vibration is proportional to kT. Thus, A is proportional to T, and λ is proportional to $1/T$. Then the resistivity given by Equation 38-14 is proportional to T, in agreement with experiment.
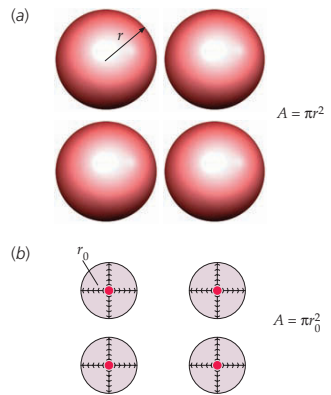
The effective area A due to thermal vibrations can be calculated, and the results give values for the resistivity that are in agreement with experiments. At $T = 300$ K, for example, the effective area turns out to be about 100 times smaller than the actual cross-sectional area of a lattice ion. We see, therefore, that the free-electron model of metals gives a good account of electrical conduction if the classical mean speed v_{av} is replaced by the Fermi speed u_F and if the collisions between electrons and the lattice ions are interpreted in terms of the scattering of electron waves, for which only deviations from a perfectly ordered lattice are important.

The presence of impurities in a metal also causes deviations from perfect regularity in the crystal lattice. The effects of impurities on resistivity are approximately independent of temperature. The resistivity of a metal containing impurities can be written $\rho = \rho_t + \rho_i$, where ρ_t is the resistivity due to the thermal motion of the lattice ions and ρ_i is the resistivity due to impurities. Figure 38-13 shows typical resistance versus temperature curves for metals with impurities. As the absolute temperature approaches zero, the resistivity due to thermal motion approaches zero, and the total resistivity approaches the resistivity due to impurities, which is constant.

FIGURE 38-12 (*a*) Classical picture of the lattice ions as spherical balls of radius r that each present an area πr^2 to the electrons. (*b*) Quantum-mechanical picture of the lattice ions as points that are vibrating in three dimensions. The area presented to the electrons is πr_0^2, where r_0 is the amplitude of oscillation of the ions.
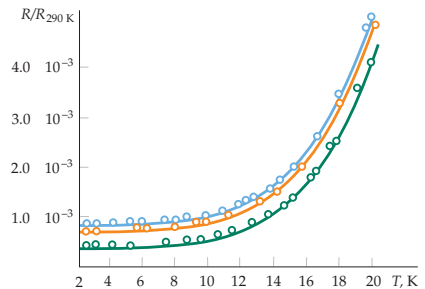
38-5 BAND THEORY OF SOLIDS

Resistivities vary enormously between insulators and conductors. For a typical insulator, such as quartz, $\rho \sim 10^{16}\ \Omega \cdot \text{m}$, whereas for a typical conductor, $\rho \sim 10^{-8}\ \Omega \cdot \text{m}$. The reason for this enormous variation is the variation in the number density of free electrons n_e. To understand this variation, we consider the effect of the lattice on the electron energy levels.

We begin by considering the energy levels of the individual atoms as they are brought together. The allowed energy levels in an isolated atom are often far apart. For example, in hydrogen, the lowest allowed energy $E_1 = -13.6$ eV is 10.2 eV below the next lowest allowed energy $E_2 = (-13.6\ \text{eV})/4 = -3.4$ eV.† Let us consider two identical atoms and focus our attention on one particular energy level. When the atoms are far apart, the energy of a particular level is the same for each atom. As the atoms are brought closer together, the energy level for each atom changes because of the influence of the other atom. As a result, the level splits into two levels of slightly different energies for the two-atom system. If we bring three atoms close together,

FIGURE 38-13 Relative resistance versus temperature for three samples of sodium. The three curves have the same temperature dependence but different magnitudes because of differing amounts of impurities in the samples.

---

* The equipartition theorem does hold for the lattice ions, which obey the Maxwell–Boltzmann energy distribution.

† The energy levels in hydrogen are discussed in Chapter 36.

a particular energy level splits into three separate levels of slightly different energies. Figure 38-14 shows the energy splitting of two energy levels for six atoms as a function of the separation of the atoms.

If we have N identical atoms, a particular energy level in the isolated atom splits into N different, closely spaced energy levels when the atoms are close together. In a macroscopic solid, N is very large—of the order of 10^{23}—so each energy level splits into a very large number of levels called a **band.** The levels are spaced almost continuously within the band. There is a separate band of levels for each particular energy level of the isolated atom. The bands may be widely separated in energy, they may be close together, or they may even overlap, depending on the kind of atom and the type of bonding in the solid.

The lowest energy bands, corresponding to the lowest energy levels of the atoms in the lattice, are filled with electrons that are bound to the individual atoms. The electrons that can take part in conduction occupy the higher energy bands. The highest energy band that contains electrons is called the **valence band.** The valence band may be completely filled with electrons or only partially filled, depending on the kind of atom and the type of bonding in the solid.

We can now understand why some solids are conductors and why others are insulators. If the valence band is only partially filled, there are many available empty energy states in the band, and the electrons in the band can easily be raised to a higher energy state by an electric field. Accordingly, this substance is a good conductor. If the valence band is filled and there is a large energy gap between it and the next available band, an applied electric field may be too weak to excite an electron from the upper energy levels of the filled band across the large gap into the energy levels of the empty band, so the substance is an insulator. The lowest band in which there are unoccupied states is called the **conduction band.** In a conductor, the valence band is only partially filled, so the valence band is also the conduction band. An energy gap between allowed bands is called a **forbidden energy band.**

The band structure for a conductor, such as copper, is shown in Figure 38-15*a*. The lower bands (not shown) are filled with the lower energy electrons of the atoms. The valence band is only about half-filled. When an electric field is established in the conductor, the electrons in the conduction band are accelerated, which means that their energies are increased. This is consistent with the exclusion principle because there are many empty energy states just above those occupied by electrons in this band. These electrons are thus the conduction electrons.

Figure 38-15*b* shows the band structure for magnesium, which is also a conductor. In this case, the highest occupied band is completely filled, but there is an empty band above it that overlaps it. The two bands thus form a combined valence–conduction band that is only partially filled.

Figure 38-15*c* shows the band structure for a typical insulator. At $T = 0$ K, the valence band is completely filled. The next energy band having empty energy states, the conduction band, is separated from the valence band by a large energy gap.
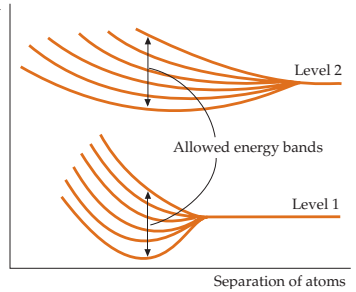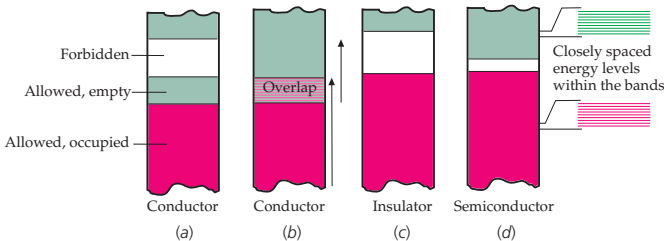
FIGURE 38-14 Energy splitting of two energy levels for six atoms as a function of the separation of the atoms. When there are many atoms, each level splits into a near-continuum of levels called a band.

FIGURE 38-15 Four possible band structures for a solid. (*a*) A typical conductor. The valence band is also the conduction band. It is only partially filled, so electrons can be easily excited to nearby energy states. (*b*) A conductor in which the valence band overlaps a conduction band above it. (*c*) A typical insulator. There is a forbidden band that has a large energy gap between the filled valence band and the conduction band. (*d*) A semiconductor. The energy gap between the filled valence band and the conduction band is very small, so some electrons are excited to the conduction band at normal temperatures, leaving holes in the valence band.

At $T = 0$, the conduction band is empty. At ordinary temperatures, a few electrons can be excited to states in that band, but most cannot be excited to states because the energy gap is large compared with the energy an electron might obtain by thermal excitation. Very few electrons can be thermally excited to the nearly empty conduction band, even at fairly high temperatures. When an electric field of ordinary magnitude is established in the solid, electrons cannot be accelerated because there are no empty energy states at nearby energies. We describe this by saying that there are no free electrons. The small conductivity that is observed is due to the very few electrons that are thermally excited into the nearly empty conduction band. When an electric field applied to an insulator is sufficiently strong to cause an electron to be excited across the energy gap to the empty band, dielectric breakdown occurs.

In some substances, the energy gap between the filled valence band and the empty conduction band is very small, as shown in Figure 38-15*d*. At $T = 0$, there are no electrons in the conduction band and the material is an insulator. At ordinary temperatures, however, there are an appreciable number of electrons in the conduction band due to thermal excitation. Such a material is called an **intrinsic semiconductor.** For typical intrinsic semiconductors, such as silicon and germanium, the energy gap is only about 1 eV. In the presence of an electric field, the electrons in the conduction band can be accelerated because there are empty states nearby. Also, for each electron in the conduction band there is a vacancy, or **hole,** in the nearly filled valence band. In the presence of an electric field, electrons in this band can also be excited to a vacant energy level. This contributes to the electric current and is most easily described as the motion of a hole in the direction of the field and opposite to the motion of the electrons. The hole thus acts like a positive charge. To visualize the conduction of holes, think of a two-lane, one-way road that has one lane completely filled with parked cars and the other lane empty. If a car moves out of the completely filled lane into the empty lane, it can move ahead freely. As the other cars move up to occupy the vacated space, the vacated space propagates backward in the direction opposite the motion of the cars. Both the forward motion of the car in the nearly empty lane and the backward propagation of the empty space contribute to a net forward propagation of the cars.

An interesting characteristic of semiconductors is that the resistivity of the substance decreases as the temperature increases, which is contrary to the case for normal conductors. The reason is that as the temperature increases, the number of free electrons increases because there are more electrons in the conduction band. The number of holes in the valence band also increases, of course. In semiconductors, the effect of the increase in the number of charge carriers, both electrons and holes, exceeds the effect of the increase in resistivity due to the increased scattering of the electrons by the lattice ions due to thermal vibrations. Semiconductors therefore have a negative temperature coefficient of resistivity.

38-6 SEMICONDUCTORS

The semiconducting property of intrinsic semiconductors makes them useful as a basis for electronic circuit components whose resistivity can be controlled by application of an external voltage or current. Most such *solid-state devices,* however, such as the semiconductor diode and the transistor, make use of **impurity semiconductors,** which are created through the controlled addition of certain impurities to intrinsic semiconductors. This process is called **doping.** Figure 38-16*a* is a schematic illustration of silicon doped with a small amount of arsenic so that the arsenic atoms replace a few of the silicon atoms in the crystal lattice. The conduction band of pure silicon is virtually empty at ordinary temperatures, so pure silicon is a poor conductor of electricity. However, arsenic has five valence electrons rather than the four valence electrons of silicon. Four of these electrons take part in bonds with the four neighboring silicon atoms, and the fifth electron is very
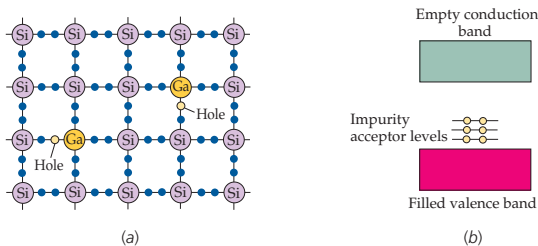
(a)

(b)
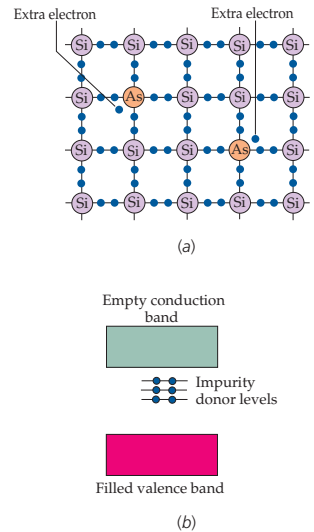
FIGURE 38-16 (*a*) A two-dimensional schematic illustration of silicon doped with arsenic. Because arsenic has five valence electrons, there is an extra, weakly bound electron that is easily excited to the conduction band, where it can contribute to electrical conduction. (*b*) Band structure of an *n*-type semiconductor, such as silicon doped with arsenic. The impurity atoms provide filled energy levels that are just below the conduction band. These levels donate electrons to the conduction band.
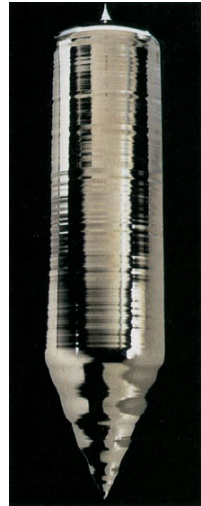
loosely bound to the atom. This extra electron occupies an energy level that is just slightly below the conduction band in the solid, and it is easily excited into the conduction band, where it can contribute to electrical conduction.

The effect on the band structure of a silicon crystal achieved by doping it with arsenic is shown in Figure 38-16b. The levels shown just below the conduction band are due to the extra electrons of the arsenic atoms. These levels are called **donor levels** because they donate electrons to the conduction band without leaving holes in the valence band. Such a semiconductor is called an **n-type semiconductor** because the major charge carriers are negatively charged electrons. The conductivity of a doped semiconductor can be controlled by controlling the amount of impurity added. The addition of just one part per million can increase the conductivity by several orders of magnitude.

Another type of impurity semiconductor can be made by replacing a silicon atom with a gallium atom, which has three valence electrons (Figure 38-17a). The gallium atom accepts electrons from the valence band to complete its four covalent bonds, thus creating a hole in the valence band. The effect on the band structure of silicon achieved by doping it with gallium is shown in Figure 38-17b. The empty levels shown just above the valence band are due to the holes from the ionized gallium atoms. These levels are called **acceptor levels** because they accept electrons from the filled valence band when those electrons are thermally excited to a higher energy state. This creates holes in the valence band that are free to propagate in the direction of an electric field. Such a semiconductor is called a **p-type semiconductor** because the charge carriers are positively charged holes. The fact that conduction is due to the motion of positively charged holes can be verified by the Hall effect. (The Hall effect is discussed in Chapter 26.)

Synthetic crystal silicon is produced beginning with a raw material containing silicon (for instance, common beach sand), separating out the silicon, and melting it. From a seed crystal, the molten silicon grows into a cylindrical crystal, such as the one shown here. The crystals (typically about 1.3 m long) are formed under highly controlled conditions to ensure that they are flawless and the crystals are then sliced into thousands of thin wafers onto which the layers of an integrated circuit are etched. *(Museum of Modern Art.)*

FIGURE 38-17 (*a*) A two-dimensional schematic illustration of silicon doped with gallium. Because gallium has only three valence electrons, there is a hole in one of its bonds. As electrons move into the hole the hole moves about, contributing to the conduction of electrical current. (*b*) Band structure of a p-type semiconductor, such as silicon doped with gallium. The impurity atoms provide empty energy levels just above the filled valence band that accept electrons from the valence band.

Example 38-7 | **Number Density of Free Electrons in Arsenic-Doped Silicon** | *Try It Yourself*

The number of free electrons in pure silicon is approximately 10^{10} electrons/cm³ at ordinary temperatures. If one silicon atom out of every 10^6 atoms is replaced by an arsenic atom, how many free electrons per cubic centimeter are there? (The density of silicon is 2.33 g/cm³ and its molar mass is 28.1 g/mol.)

PICTURE The number of silicon atoms per cubic centimeter, n_{Si}, can be found from $n_{Si} = \rho N_A / M$. Then, because each arsenic atom contributes one free electron, the number of electrons contributed by the arsenic atoms is $10^{-6} n_{Si}$.

SOLVE
Cover the column to the right and try these on your own before looking at the answers.

Steps	Answers
1. Calculate the number of silicon atoms per cubic centimeter.	$n_{Si} = \dfrac{\rho N_A}{M}$ $= \dfrac{(2.33 \text{ g/cm}^3)(6.02 \times 10^{23} \text{ atoms/mol})}{28.1 \text{ g/mol}}$ $= 4.99 \times 10^{22} \text{ atoms/cm}^3$
2. Multiply by 10^{-6} to obtain the number of arsenic atoms per cubic centimeter, which equals the added number of free electrons per cubic centimeter.	$n_{As} = 10^{-6} n_{Si} = 4.99 \times 10^{16} \text{ atoms/cm}^3$
3. The number of free electrons per cubic centimeter is equal to the number of arsenic atoms per cubic centimeter plus 1×10^{-10} (the number of silicon atoms per cubic centimeter).	$n_e = n_{As} + 1 \times 10^{-10} n_{Si}$ $= 4.99 \times 10^{16} \text{ cm}^{-3} + 1 \times 10^{10} \text{ cm}^{-3}$ $\approx \boxed{5 \times 10^{16} \text{ electrons/cm}^3}$

CHECK As expected, the step-3 result is less than the number density of silicon atoms and more than the number density of conduction electrons in pure silicon.
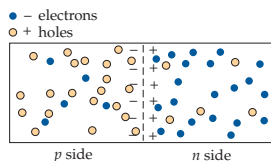
TAKING IT FURTHER Because silicon has so few free electrons per atom, the number density of conduction electrons is increased by a factor of approximately 5 million per cubic centimeter by doping silicon with just one arsenic atom per million silicon atoms.

PRACTICE PROBLEM 38-3 How many free electrons are there per silicon atom in pure silicon?
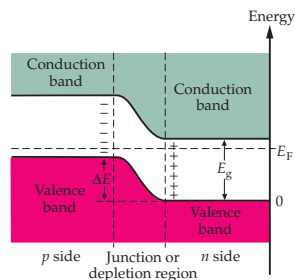

● − electrons
○ + holes

p side   *n* side

FIGURE 38-18 A *pn* junction. Because of the difference in their concentrations on either side of the *pn* junction, holes diffuse from the *p* side to the *n* side, and electrons diffuse from the *n* side to the *p* side. As a result, there is a double layer of charge at the junction, with the *p* side being negative and the *n* side being positive.

★ 38-7 SEMICONDUCTOR JUNCTIONS AND DEVICES

Semiconductor devices such as diodes and transistors make use of *n*-type semiconductors and *p*-type semiconductors joined together, as shown in Figure 38-18. In practice, the two types of semiconductors are often incorporated into a single silicon crystal doped with donor impurities on one side and acceptor impurities on the other side. The region in which the semiconductor changes from a *p*-type semiconductor to an *n*-type semiconductor is called a **pn junction.**
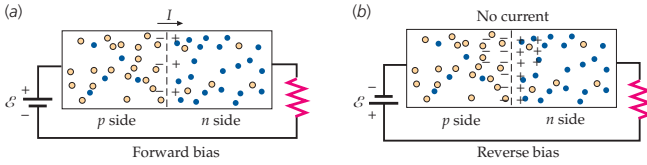
When an *n*-type semiconductor and a *p*-type semiconductor are placed in contact, the initially unequal concentrations of electrons and holes result in the diffusion of electrons across the junction from the *n* side to the *p* side and holes from the *p* side to the *n* side until equilibrium is established. The result of this diffusion is a net transport of positive charge from the *p* side to the *n* side. Unlike the case when two different metals are in contact, the electrons cannot travel very far from the junction region because the semiconductor is not a particularly good conductor. The diffusion of electrons and holes therefore creates a double layer of charge at the junction similar to that on a parallel-plate capacitor. There is, thus, a potential difference *V* across the junction, which tends to inhibit further diffusion. In equilibrium, the *n* side which has a net positive charge will be at a higher potential than the *p* side which has a net negative charge. In the junction region, between the charge layers, there will be very few charge carriers of either type, so the junction region has a high resistance. Figure 38-19 shows the energy-level diagram for a *pn* junction. The junction region is also called the **depletion region** because it has been depleted of charge carriers.

Energy

Conduction band

Conduction band

E_F

Valence band ΔE E_g

Valence band

0

p side   Junction or depletion region   *n* side

FIGURE 38-19 Electron energy levels for a *pn* junction.

(a)

(b) No current

Forward bias

Reverse bias

FIGURE 38-20 A *pn*-junction diode. (*a*) Forward-biased *pn* junction. The applied potential difference enhances the diffusion of holes from the *p* side to the *n* side and of electrons from the *n* side to the *p* side, resulting in a current *I*. (*b*) Reverse-biased *pn* junction. The applied potential difference inhibits the further diffusion of holes and electrons across the junction, so there is no current.
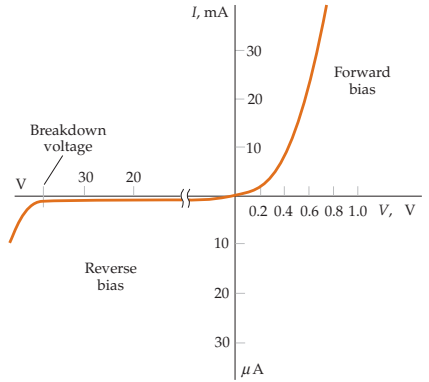
*DIODES

In Figure 38-20, an external potential difference has been applied across a *pn* junction by connecting a battery and a resistor to the semiconductor. When the positive terminal of the battery is connected to the *p* side of the junction, as shown in Figure 38-20*a,* the junction is said to be **forward biased.** Forward biasing lowers the potential across the junction. The diffusion of electrons and holes is thereby increased as they attempt to reestablish equilibrium, resulting in a current in the circuit.

If the positive terminal of the battery is connected to the *n* side of the junction, as shown in Figure 38-20*b,* the junction is said to be **reverse biased.** Reverse biasing tends to increase the potential difference across the junction, thereby further inhibiting diffusion. Figure 38-21 shows a plot of current versus voltage for a typical semiconductor junction. Essentially, the junction conducts only in one direction for applied voltages greater than the breakdown voltage. A single-junction semiconductor device is called a **diode.*** Diodes have many uses. One use is to convert alternating current into direct current, a process called *rectification.*

Note that the current in Figure 38-21 suddenly increases in magnitude at extreme values of reverse bias. In such large electric fields, electrons are stripped from their atomic bonds and accelerated across the junction. These electrons, in turn, cause others to break loose. This effect is called **avalanche breakdown.** Although such a breakdown can be disastrous in a circuit where it is not intended, the fact that it occurs at a sharply defined voltage makes it of use in a special voltage reference standard known as a **Zener diode.** Zener diodes are also used to protect devices from excessively high voltages.

An interesting effect, one that we discuss only qualitatively, occurs if both the *n* side and the *p* side of a *pn*-junction diode are so heavily doped that the donors on the *n* side provide so many electrons that the lower part of the conduction band is practically filled, and the acceptors on the *p* side accept so many electrons that the upper part of the valence band is nearly empty. Figure 38-22*a* shows the energy-level

FIGURE 38-21 Plot of current versus applied voltage across a *pn* junction. Note the different scales on both axes for the forward and reverse bias conditions.

FIGURE 38-22 Electron energy levels for a heavily doped *pn*-junction tunnel diode. (*a*) With no bias voltage, some electrons tunnel in each direction. (*b*) With a small bias voltage, the tunneling current is enhanced in one direction, making a sizable contribution to the net current. (*c*) With further increases in the bias voltage, the tunneling current decreases dramatically.

(a)

(b)

(c)
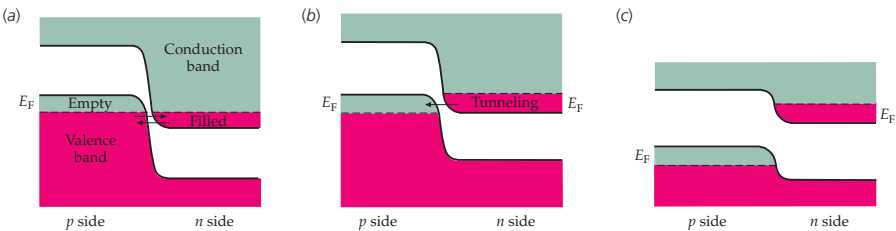
* The name *diode* originates from a vacuum tube device consisting of just two electrodes that also conducts electric current in one direction only.

diagram for this situation. Because the depletion region is now so narrow, electrons can easily penetrate the potential barrier across the junction and tunnel to the other side. The flow of electrons through the barrier is called a **tunneling current,** and such a heavily doped diode is called a **tunnel diode.**

At equilibrium where there is no bias, there is an equal tunneling current in each direction. When a small bias voltage is applied across the junction, the energy-level diagram is as shown in Figure 38-22b, and the tunneling of electrons from the n side to the p side is increased, whereas the tunneling of electrons in the opposite direction is decreased. This tunneling current, in addition to the usual current due to diffusion, results in a considerable net current. When the bias voltage is increased slightly, the energy-level diagram is as shown in Figure 38-22c, and the tunneling current is decreased. Although the diffusion current is increased, the net current is decreased. At large bias voltages, the tunneling current is completely negligible, and the total current increases with increasing bias voltage due to diffusion, as in an ordinary pn-junction diode. Figure 38-23 shows the current versus voltage curve for a tunnel diode. Such diodes are used in electric circuits because of their very fast response time. When operated near the peak in the current versus voltage curve, a small change in bias voltage results in a large change in the current.

Another use for the pn-junction semiconductor is the **solar cell,** which is illustrated schematically in Figure 38-24. When a photon of energy greater than the gap energy (1.1 eV in silicon) strikes the p-type region, it can excite an electron from the valence band into the conduction band, leaving a hole in the valence band. This region is already rich in holes. Some of the electrons created by the photons will recombine with holes, but some will migrate to the junction. From there, they are accelerated into the n-type region by the electric field between the double layer of charge. This creates an excess negative charge in the n-type region and an excess positive charge in the p-type region. The result is a potential difference between the two regions, which in practice is approximately 0.6 V. If a load resistance is connected across the two regions, a charge flows through the resistance. Some of the incident light energy is thus converted into electrical energy. The current in the resistor is proportional to the rate of arrival of incident photons, which is in turn proportional to the intensity of the incident light.

There are many other applications of semiconductors with pn junctions. Particle detectors, called **surface-barrier detectors,** consist of a pn-junction semiconductor that has a large reverse bias so that there is ordinarily no current. When a high-energy particle, such as an electron, passes through the semiconductor, it creates many electron–hole pairs as it loses energy. The resulting current pulse signals the passage of the particle. **Light-emitting diodes** (LEDs) are pn-junction semiconductors that have large forward biases that produce large excess concentrations of electrons on the p sides and holes on the n sides of the junctions. Under these conditions, an LED emits light as the electrons and holes recombine. This is essentially the reverse of the process that occurs in a solar cell, in which electron–hole pairs are created by the absorption of light. LEDs are commonly used as warning indicators and as sources of infrared light beams.

*TRANSISTORS

The transistor, a semiconducting device that is used to produce a desired output signal in response to an input signal, was invented in 1948 by William Shockley, John Bardeen, and Walter Brattain and has revolutionized the electronics industry and our everyday world. A *simple bipolar junction transistor** consists of three distinct semiconductor regions called the **emitter,** the **base,** and the **collector.** The base is a very thin region of one type of semiconductor sandwiched between two regions of the opposite type. The emitter semiconductor is much more heavily

---

* Besides the bipolar junction transistor, there are other categories of transistors, notably, the field–effect transistor.
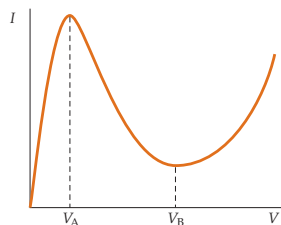
FIGURE 38-23 Current versus applied (bias) voltage V for a tunnel diode. For $V < V_A$, an increase in the bias voltage V enhances tunneling. For $V_A < V < V_B$, an increase in the bias voltage inhibits tunneling. For $V > V_B$, the tunneling is negligible, and the diode behaves like an ordinary pn-junction diode.
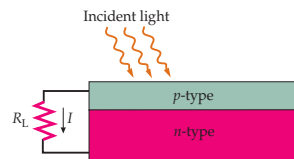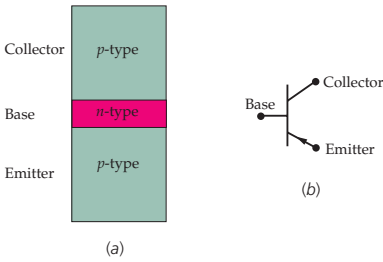

FIGURE 38-24 A pn-junction semiconductor as a solar cell. When light strikes the p-type region, electron-hole pairs are created, resulting in a current through the load resistance R_L.
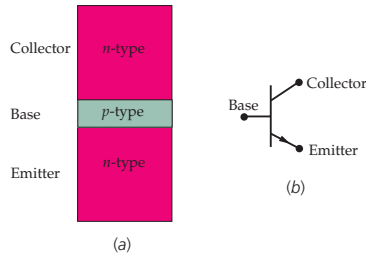

A light-emitting diode (LED). (© C. Falco/Photo Researchers.)

FIGURE 38-25 A *pnp* transistor. (*a*) The heavily doped emitter emits holes that pass through the thin base to the collector. (*b*) Symbol for a *pnp* transistor in a circuit. The arrow points in the direction of the conventional current, which is the same as that of the emitted holes.

FIGURE 38-26 An *npn* transistor. (*a*) The heavily doped emitter emits electrons that pass through the thin base to the collector. (*b*) Symbol for an *npn* transistor. The arrow points in the direction of the conventional current, which is opposite the direction of the emitted electrons.

doped than either the base or the collector. In an *npn* transistor, the emitter and collector are *n*-type semiconductors and the base is a *p*-type semiconductor; in a *pnp* transistor, the base is an *n*-type semiconductor and the emitter and collector are *p*-type semiconductors.

Figure 38-25 and Figure 38-26 show, respectively, a *npn* transistor and an *npn* transistor, along with the symbols used to represent each transistor in circuit diagrams. We see that either transistor consists of two *pn* junctions. We will discuss the operation of a *npn* transistor. The operation of an *npn* transistor is similar.
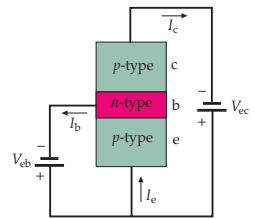
In the normal operation of a *pnp* transistor, the emitter-base junction is forward biased, and the base-collector junction is reverse biased, as shown in Figure 38-27. The heavily doped *p*-type emitter emits holes that flow toward the emitter-base junction. This flow constitutes the emitter current I_e. Because the base is very thin, most of the holes flow across the base into the collector. This flow in the collector constitutes a current I_c. However, some of the holes recombine in the base producing a positive charge that inhibits the further flow of charge. To prevent this, some of the holes that do not reach the collector are drawn off the base as a base current I_b in a wire connected to the base. In Figure 38-27, therefore, I_c is almost but not quite equal to I_e, and I_b is much smaller than either I_c or I_e. It is customary to express I_c as

$$I_c = \beta I_b \qquad\qquad 38\text{-}31$$

where β is called the current gain of the transistor. Transistors can be designed to have values of β as low as ten or as high as several hundred.

Figure 38-28 shows a simple *pnp* transistor used as an amplifier. A small, time-varying input voltage v_s is connected in series with a constant bias voltage V_{eb}. The base current is then the sum of a steady current I_b produced by the bias voltage V_{eb} and a time-varying current i_b due to the signal voltage v_s. Because v_s may at any instant be either positive or negative, the bias voltage V_{eb} must be large enough to ensure that there is always a forward bias on the emitter-base junction. The collector current will consist of two parts: a constant direct current $I_c = \beta I_b$ and a time-varyng current $i_c = \beta i_b$. We thus have a current amplifier in which the time-varying output current i_c is β multiplied by the input current i_b. In such an amplifier, the steady currents I_c and I_b, although essential to the operation of the transistor, are usually not of interest. The input signal voltage v_s is related to the base current by Ohm's law:

$$i_b = \frac{v_s}{R_b + r_b} \qquad\qquad 38\text{-}32$$

FIGURE 38-27 A *pnp* transistor biased for normal operation. Holes from the emitter can easily diffuse across the base, which is only tens of nanometers thick. Most of the holes flow to the collector, producing current I_c.
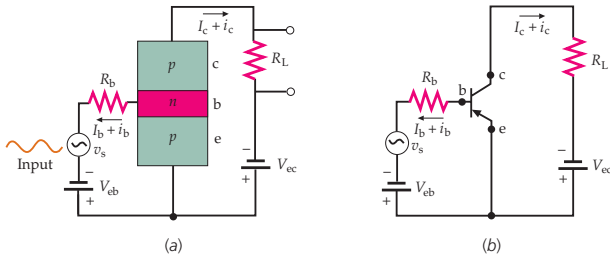
FIGURE 38-28 (a) A *pnp* transistor used as an amplifier. A small change i_b in the base current results in a large change i_c in the collector current. Thus, a small signal in the base circuit results in a large signal across the load resistor R_L in the collector circuit. (b) The same circuit as in Figure 38-28a with the conventional symbol for the transistor.

where r_b is the internal resistance of that part of the transistor between the base and emitter. Similarly, the collector current i_c produces a time-varying voltage v_L across the output or load resistance R_L given by

$$v_L = i_c R_L \qquad\qquad 38\text{-}33$$

Using Equation 38-31 and Equation 38-32, we have

$$i_c = \beta i_b = \beta \frac{v_s}{R_b + r_b} \qquad\qquad 38\text{-}34$$

The output voltage is thus related to the input voltage by

$$v_L = \beta \frac{v_s}{R_b + r_b} R_L = \beta \frac{R_L}{R_b + r_b} v_s \qquad\qquad 38\text{-}35$$

The ratio of the output voltage to the input voltage is the **voltage gain** of the amplifier:

$$\text{Voltage gain} = \frac{v_L}{v_s} = \beta \frac{R_L}{R_b + r_b} \qquad\qquad 38\text{-}36$$

A typical amplifier (for example, in a tape player) has several transistors, similar to the one shown in Figure 38-28, connected in series so that the output of one transistor serves as the input for the next. Thus, the very small voltage fluctuations produced by the motion of the magnetic tape past the pickup heads controls the large amounts of power required to drive the loudspeakers. The power delivered to the speakers is supplied by the dc sources connected to each transistor.

The technology of semiconductors extends well beyond individual transistors and diodes. Many of the electronic devices we use every day, such as laptop computers and the processors that govern the operation of vehicles and appliances, rely on large-scale integration of many transistors and other circuit components on a single chip. Large-scale integration combined with advanced concepts in semiconductor theory has created remarkable new instruments for scientific research.

38-8 SUPERCONDUCTIVITY

There are some substances for which the resistivity suddenly drops to zero below a certain temperature T_c, which is called the **critical temperature.** This amazing phenomenon, called **superconductivity,** was discovered in 1911 by the Dutch physicist H. Kamerlingh Onnes, who developed a technique for liquefying

helium (boiling point equal to 4.2 K) and used his technique to explore the properties of substances at temperatures in that range. Figure 38-29 shows Onnes's plot of the resistance of mercury versus temperature. The critical temperature for mercury is approximately the same as the boiling point of helium, which is 4.2 K. Critical temperatures for other superconducting elements range from less than 0.1 K for hafnium and iridium to 9.2 K for niobium. The temperature range for superconductors is much higher for a number of metallic compounds. For example, the superconducting alloy Nb_3Ge, discovered in 1973, has a critical temperature of 25 K, which was the highest known until 1986, when the discoveries of J. Georg Bednorz and K. Alexander Müller launched the era of high-temperature superconductors, now defined as materials that exhibit superconductivity at temperatures above 77 K (the temperature at which nitrogen boils). The highest temperature at which superconductivity has been demonstrated, using thallium-doped $HgBa_2Ca_2Cu_3O_8$ + delta, is 138 K at atmospheric pressure. At extremely high pressures, some materials exhibit superconductivity at temperatures as high as 164 K.

The resistivity of a superconductor is zero. There can be a current in a superconductor even when there is no emf in the superconducting circuit. Indeed, in superconducting rings in which there was no electric field, steady currents have been observed to persist for years without apparent loss. Despite the cost and inconvenience of refrigeration using expensive liquid helium, many superconducting magnets have been built using superconducting materials, because such magnets require no power expenditure to maintain the large current needed to produce a large magnetic field.
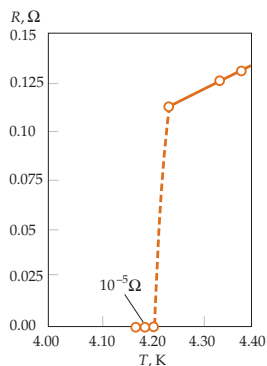
FIGURE 38-29 Plot by H. Kamerlingh Onnes of the resistance of mercury versus temperature, showing the sudden decrease at the critical temperature of T = 4.2 K.

The wires for the magnetic field of a magnetic resonance imaging (MRI) machine carry large currents. To keep the wires from overheating, they are maintained at superconducting temperatures. To accomplish this, they are immersed in liquid helium. *(Corbis.)*

The discovery of high-temperature superconductors has revolutionized the study of superconductivity because relatively inexpensive liquid nitrogen, which boils at 77 K, can be used for a coolant. However, many problems, such as brittleness and the toxicity of the materials, make these new superconductors difficult to use. The search continues for new materials that will be superconductors at even higher temperatures.

THE BCS THEORY

It had been recognized for some time that low temperature superconductivity is due to a collective action of the conducting electrons. In 1957, John Bardeen, Leon Cooper, and Robert Schrieffer published a successful theory of low temperature superconductivity now known by the initials of the inventors as the **BCS theory.** According to this theory, the electrons in a superconductor are coupled in pairs at low temperatures. The coupling comes about because of the interaction between electrons and the crystal lattice. One electron interacts with the lattice and perturbs it. The perturbed lattice interacts with another electron in such a way that there is an attraction between the two electrons that at low temperatures can exceed the Coulomb repulsion between them. The electrons form a bound state called a **Cooper pair.** The electrons in a Cooper pair have equal and opposite spins, so they form a system with zero spin. Each Cooper pair acts as a *single particle* with zero spin, in other words, as a boson. Bosons do not obey the exclusion principle. Any number of Cooper pairs may be in the same quantum state with the same energy. In the ground state of a superconductor (at $T = 0$), all the conduction electrons are in Cooper pairs and all the Cooper pairs are in the same energy state. In the superconducting state, the Cooper pairs are correlated so that they act collectively. An electric current can be produced in a superconductor because all of the electrons in this collective state move together. But energy cannot be dissipated by individual collisions of electron and lattice ions unless the temperature is high enough to break the binding of the Cooper pairs. The required energy is called the *superconducting energy gap* E_g. In the BCS theory, this energy at zero temperature is related to the critical temperature by

$$E_g = \tfrac{7}{2}kT_c \qquad\qquad 38\text{-}37$$

The energy gap can be determined by measuring the current across a junction between a normal metal and a superconductor as a function of voltage. Consider two metals separated by a layer of insulating material, such as aluminum oxide, that is only a few nanometers thick. The insulating material between the metals forms a barrier that prevents most electrons from traversing the junction. However, waves can tunnel through a barrier if the barrier is not too thick, even if the energy of the wave is less than that of the barrier.

When the materials on either side of the gap are normal nonsuperconducting metals, the current resulting from the tunneling of electrons through the insulating layer obeys Ohm's law for low applied voltages (Figure 38-30a). When one of the metals is a normal metal and the other is a superconductor, there is no current (at absolute zero) unless the applied voltage V is greater than a critical voltage $V_c = E_g/(2e)$, where E_g is the superconductor energy gap. Figure 38-30b shows the plot of current versus voltage for this situation. The current escalates rapidly when the energy $2eV$ absorbed by a Cooper pair traversing the barrier approaches $E_g = 2eV_c$, the minimum energy needed to break up the pair. (The small current visible in Figure 38-30b before the critical voltage is reached is present because at any temperature above absolute zero some of the electrons in the superconductor are thermally excited above the energy gap and are therefore not paired.) At voltages slightly above V_c, the current versus voltage curve becomes that for a normal metal. The superconducting energy gap can thus be measured by measuring the average voltage for the transition region.
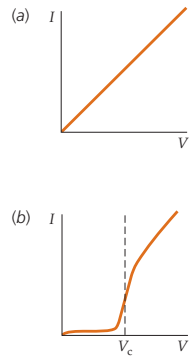
FIGURE 38-30 Tunneling current versus voltage for a junction of two metals separated by a thin oxide layer. (*a*) When both metals are normal metals, the current is proportional to the voltage, as predicted by Ohm's law. (*b*) When one metal is a normal metal and another metal is a superconductor, the current is approximately zero until the applied voltage V approaches the critical voltage $V_c = E_g/(2e)$.

Example 38-8	**Superconducting Energy Gap for Mercury**

Calculate the superconducting energy gap for mercury (T_c = 4.2 K) predicted by the BCS theory.

PICTURE The energy gap is related to the critical temperature by E_g = 3.5 kT_c (Equation 38-37).

SOLVE

1. The BCS prediction for the energy gap is $\qquad$ $E_g = 3.5kT_c$

2. Substitute T_c = 4.2 K: $\qquad$ $E_g = 3.5kT_c$

$$= 3.5(1.38 \times 10^{-23} \text{ J/K})(4.2 \text{ K})\left(\frac{1 \text{ ev}}{1.6 \times 10^{-19} \text{ J}}\right)$$

$$= \boxed{1.3 \times 10^{-3} \text{ eV}}$$

Note that the energy gap for a typical superconductor is much smaller than the energy gap for a typical semiconductor, which is of the order of 1 eV. As the temperature is increased from $T = 0$, some of the Cooper pairs are broken. Then there are fewer pairs available for each pair to interact with, and the energy gap is reduced until at $T = T_c$ the energy gap is zero (Figure 38-31).

THE JOSEPHSON EFFECT

When two superconductors are separated by a thin nonsupercon-ducting barrier (for example, a layer of aluminum oxide a few nanometers thick), the junction is called a **Josephson junction,** based on the prediction in 1962 by Brian Josephson that Cooper pairs could tunnel across such a junction from one superconductor to the other with no resistance. The tunneling of Cooper pairs constitutes a current, which does not require a voltage to be applied across the junction. The current depends on the difference in phase of the wave functions that describe the Cooper pairs. Let ϕ_1 be the phase constant for the wave function of a Cooper pair in one superconductor. All the Cooper pairs in a superconductor act coherently and have the same phase constant. If ϕ_2 is the phase constant for the Cooper pairs in the second superconductor, the current across the junction is given by

$$I = I_{max} \sin(\phi_2 - \phi_1) \qquad \text{38-38}$$

where I_{max} is the maximum current, which depends on the thickness of the barrier. This result has been observed experimentally and is known as the **dc Josephson effect.**

Josephson also predicted that if a dc voltage V were applied across a Josephson junction, there would be a current that alternates with frequency f given by

$$f = \frac{2e}{h}V \qquad \text{38-39}$$

This result, known as the **ac Josephson effect,** has been observed experimentally, and careful measurement of the frequency allows a precise determination of the ratio e/h. Because frequency can be measured very accurately, the ac Josephson effect is also used to establish precise voltage standards. The inverse effect, in which the application of an alternating voltage across a Josephson junction results in a dc current, has also been observed.
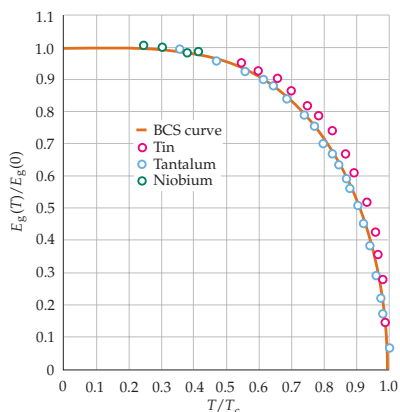
FIGURE 38-31 Ratio of the energy gap at temperature T to that at temperature $T = 0$ as a function of the relative temperature T/T_c. The solid curve is that predicted by the BCS theory.

Example 38-9   **Frequency of Josephson Current**

Using $e = 1.602 \times 10^{-19}$ C and $h = 6.626 \times 10^{-34}$ J·s, calculate the frequency of the Josephson current if the applied voltage is 1.000 μV.

PICTURE The frequency f is related to the applied voltage V by $hf = 2eV$ (Equation 38-39).

SOLVE

Substitute the given values into Equation 38-39 to calculate f:

$$f = \frac{2e}{h} V = \frac{2(1.602 \times 10^{-19}\text{ C})}{6.626 \times 10^{-34}\text{ J·s}}(1.000 \times 10^{-6}\text{ V})$$

$$= 4.835 \times 10^{8}\text{ Hz} = \boxed{483.5\text{ MHz}}$$

38-9 THE FERMI–DIRAC DISTRIBUTION

The classical Maxwell–Boltzmann distribution (Equation 17-38) gives the number dN of molecules that have energy E in the range between E and $E + dE$. The number dN is equal to the product of $g(E)\,dE$ where $g(E)$ is the **density of states** (number of energy states in the range dE) and the Boltzmann factor $e^{-E/(kT)}$, which is the probability of a state being occupied. The distribution function for free electrons in a metal is called the **Fermi–Dirac distribution.** The Fermi–Dirac distribution can be written in the same form as the Maxwell–Boltzmann distribution, where the density of states calculated from quantum theory and the Boltzmann factor is replaced by the Fermi factor. Let $n(E)\,dE$ be the number of electrons that have energies between E and $E + dE$. This number is written

$$n(E)dE = f(E)g(E)dE \qquad\qquad 38\text{-}40$$

ENERGY DISTRIBUTION FUNCTION

where $g(E)\,dE$ is the number of states that have energies between E and $E + dE$ and $f(E)$ is the probability of a state being occupied, which is the Fermi factor. The density of states in three dimensions is somewhat challenging to calculate, so we just give the result. For electrons in a metal of volume V, the density of states is

$$g(E) = \frac{8\sqrt{2}\pi m_e^{3/2} V}{h^3} E^{1/2} \qquad\qquad 38\text{-}41$$

DENSITY OF STATES

As in the classical Maxwell–Boltzmann distribution, the density of states is proportional to $E^{1/2}$.

At $T = 0$, the Fermi factor is given by Equation 38-24:

$$f(E) = \begin{cases} 1 & E < E_F \\ 0 & E > E_F \end{cases}$$

The integral of $n(E)\,dE$ over all energies gives the total number of electrons N. We can derive the equation

$$E_F = \frac{h^2}{8m_e}\left(\frac{3N}{\pi V}\right)^{2/3}$$

(Equation 38-22a) for the Fermi energy at $T = 0$ by integrating $n(E) dE$ from $E = 0$ to $E = \infty$. We obtain

$$N = \int_0^\infty n(E)dE = \int_0^{E_F} n(E)dE + \int_{E_F}^\infty n(E)dE$$

$$= \int_0^{E_F} \frac{8\sqrt{2}\pi m_e^{3/2}V}{h^3} E^{1/2} \, dE + 0 = \frac{16\sqrt{2}\pi m_e^{3/2}V}{3h^3} E_F^{3/2}$$

Note that at $T = 0$, $n(E)$ is zero for $E > E_F$. Solving for E_F gives the Fermi energy at $T = 0$:

$$E_F = \frac{h^2}{8m_e}\left(\frac{3N}{\pi V}\right)^{2/3} \qquad\qquad 38\text{-}42$$

which is Equation 38-22a. In terms of the Fermi energy, the density of states (Equation 38-41) is

$$g(E) = \frac{8\sqrt{2}\pi m_e^{3/2}V}{h^3} E^{1/2} = \frac{3}{2}NE_F^{-3/2}E^{1/2} \qquad\qquad 38\text{-}43$$

DENSITY OF STATES IN TERMS OF E_F

which is obtained by solving Equation 38-42 for m_e, and then substituting for m_e in Equation 38-41. The average energy at $T = 0$ is calculated from

$$E_{av} = \frac{\displaystyle\int_0^{E_F} Eg(E)dE}{\displaystyle\int_0^{E_F} g(E)dE} = \frac{1}{N}\int_0^{E_F} Eg(E)dE \qquad\qquad 38\text{-}44$$

where $N = \displaystyle\int_0^{E_F} g(E)dE$ is the total number of electrons. Substituting for $g(E)$ from Equation 38-43 and then evaluating the integral in Equation 38-44, we obtain Equation 38-23:

$$E_{av} = \tfrac{3}{5}E_F \qquad\qquad 38\text{-}45$$

AVERAGE ENERGY AT $T = 0$

At $T > 0$, the Fermi factor is more complicated. It can be shown to be

$$f(E) = \frac{1}{e^{(E-E_F)/(kT)} + 1} \qquad\qquad 38\text{-}46$$

FERMI FACTOR

We can see from this equation that for E greater than E_F, $e^{(E-E_F)/(kT)}$ becomes very large as T approaches zero, so at $T = 0$, the Fermi factor is zero for $E > E_F$. On the other hand, for E less than E_F, $e^{(E-E_F)/(kT)}$ approaches 0 as T approaches zero, so at $T = 0$, $f(E) = 1$ for $E < E_F$. Thus, the Fermi factor given by Equation 38-46 holds for all temperatures. Note also that for any nonzero value of T, $f(E) = \tfrac{1}{2}$ at $E = E_F$.

The complete Fermi-Dirac distribution function is thus

$$n(E)dE = g(E)f(E)dE = \frac{8\sqrt{2}\pi m_e^{3/2}V}{h^3}E^{1/2}\frac{1}{e^{(E-E_F)/(kT)}+1}dE \qquad 38\text{-}47$$

<div align="center">FERMI–DIRAC DISTRIBUTION</div>

We can see that for those few electrons that have energies much greater than the Fermi energy, the Fermi factor approaches $1/e^{(E-E_F)/(kT)} = e^{(E_F-E)/(kT)} = e^{E_F/(kT)}e^{-E/(kT)}$, which is proportional to $e^{-E/(kT)}$. Thus, the high-energy tail of the Fermi–Dirac energy distribution decreases with increasing E as $e^{-E/(kT)}$, just like the classical Maxwell–Boltzmann energy distribution. The reason for this is that in this high-energy region there are many unoccupied energy states and few electrons, so the exclusion principle is not important. Thus, the Fermi–Dirac distribution approaches the classical Maxwell–Boltzmann distribution in the high-energy limit. This result has practical importance because it applies to the conduction electrons in semiconductors.

Example 38-10   Fermi Factor for Copper at 300 K

At what energy is the Fermi factor equal to 0.100 for copper at $T = 300$ K?

PICTURE We set $f(E) = 0.100$ in Equation 38-46, using $T = 300$ K and $E_F = 7.03$ eV from Table 38-1, and solve for E.

SOLVE

1. Solve Equation 38-46 for $e^{(E-E_F)/(kT)}$:

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

so

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

2. Take the logarithm of both sides:

$$\frac{E-E_F}{kT} = \ln\left[\frac{1}{f(E)}-1\right]$$

3. Solve for E. For E_F, use the value for E_F at $T = 0$ K listed in Table 38-1:

$$E = E_F + \left[\frac{1}{f(E)}-1\right]kT$$

$$= 7.03\text{ eV} + \ln\left[\frac{1}{0.100}-1\right](8.62\times10^{-5}\text{ eV/K})(300\text{ K})$$

$$= \boxed{7.09\text{ eV}}$$

CHECK As expected, the energy is slightly above the Fermi energy when the Fermi factor is equal to 0.100.

Example 38-11   Probability of a Higher Energy State Being Occupied

Find the probability that an energy state in copper 0.100 eV above the Fermi energy is occupied at $T = 300$ K.

PICTURE The probability is the Fermi factor given in Equation 38-46, with $E_F = 7.03$ eV and $E = 7.13$ eV.

SOLVE

1. The probability of an energy state being occupied equals the Fermi factor:

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

2. Calculate the exponent in the Fermi factor (exponents are always dimensionless):

$$\frac{E - E_F}{kT} = \frac{7.13\ \text{eV} - 7.03\ \text{eV}}{(8.62 \times 10^{-5}\ \text{eV/K})(300\ \text{K})} = 3.87$$

3. Use this result to calculate the Fermi factor:

$$f(E) = \frac{1}{e^{(E-E_F)/(kT)} + 1} = \frac{1}{e^{3.87} + 1}$$

$$= \frac{1}{48 + 1} = \boxed{0.020}$$

CHECK The probability that an energy state above the Fermi energy is occupied is less than one-half. As expected, the step 4 result less than one-half.

TAKING IT FURTHER The probability of an electron having an energy 0.100 eV above the Fermi energy at 300 K is only about 2 percent.

Example 38-12 **Probability of a Lower Energy State Being Occupied**   *Try It Yourself*

Find the probability that an energy state in copper 0.10 eV *below* the Fermi energy is occupied at $T = 300\ \text{K}$.

PICTURE The probability is the Fermi factor given in Equation 38-46, with $E_F = 7.03\ \text{eV}$ and $E = 6.93\ \text{eV}$.

SOLVE
Cover the column to the right and try these on your own before looking at the answers.

Steps	Answers
1. Write the Fermi factor:	$f(E) = \dfrac{1}{e^{(E-E_F)/(kT)} + 1}$
2. Calculate the exponent in the Fermi factor:	$\dfrac{E - E_F}{kT} = \dfrac{6.93\ \text{eV} - 7.03\ \text{eV}}{(8.62 \times 10^{-5}\ \text{eV/K})(300\ \text{K})} = -3.87$
3. Use your result from step 2 to calculate the Fermi factor:	$f(E) = \dfrac{1}{e^{(E-E_F)/(kT)} + 1} = \dfrac{1}{e^{3.87} + 1}$ $= \dfrac{1}{0.021 + 1} = \boxed{0.98}$

CHECK As expected, the step-3 result is greater than one-half.

TAKING IT FURTHER The probability of an electron having an energy of 0.10 eV *below* the Fermi energy at 300 K is approximately 98 percent.

PRACTICE PROBLEM 38-4 What is the probability of an energy state 0.10 eV below the Fermi energy being unoccupied at 300 K?