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, 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

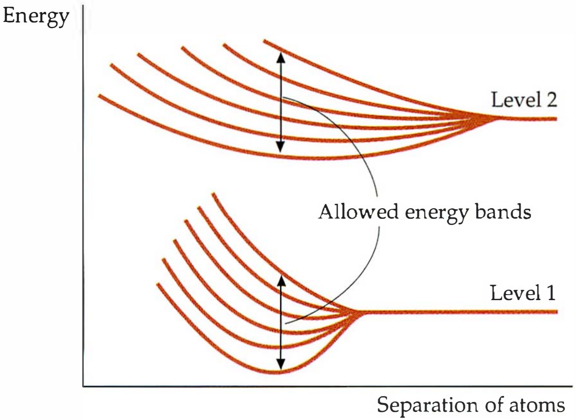


FIGURE 38-14 Energy splitting of two energy levels for six atoms as a function of the separation of the atoms. When there are many atoms, each level splits into a near-continuum of levels called a band.

---

† 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.

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.

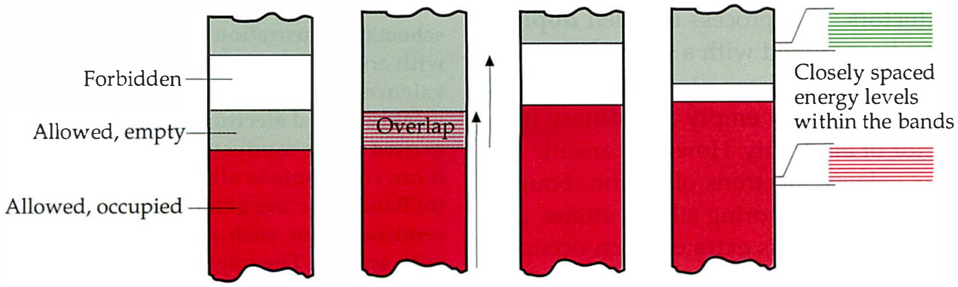
The lowest energy bands, corresponding to the lowest energy levels of the atoms in the lattice, are filled with electrons that are bound to the individual atoms. The electrons that can take part in conduction occupy the higher energy bands. The highest energy band that contains electrons is called the **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 full, 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 material is a good conductor. If the valence band is full and there is a large energy gap between it and the next available band, a typical applied electric field will 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 material 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 inner electrons of the atoms. The valence band is only about half full. When an electric field is established in the conductor, the electrons in the conduction band are accelerated, which means that their energy is increased. This is consistent with the Pauli 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 full, 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 full. The next energy band containing empty energy states, the conduction band, is separated from the valence band by a large energy gap. At $T = 0$, the conduction band is empty. At ordinary temperatures, a few electrons can be excited to states in this 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



Conductor (*a*) Conductor (*b*) Insulator (*c*) Semiconductor (*d*)

FIGURE 38-15 Four possible band structures for a solid. (*a*) A typical conductor. The valence band is only partially full, so electrons can be easily excited to nearby energy states. (*b*) A conductor in which the allowed energy bands overlap. (*c*) A typical insulator. There is a forbidden band with 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.

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 materials, 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. However, at ordinary temperatures, 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 with one lane full of parked cars and the other lane empty. If a car moves out of the filled lane into the empty lane, it can move ahead freely. As the other cars move up to occupy the space left, the empty 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 material 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 materials 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 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.

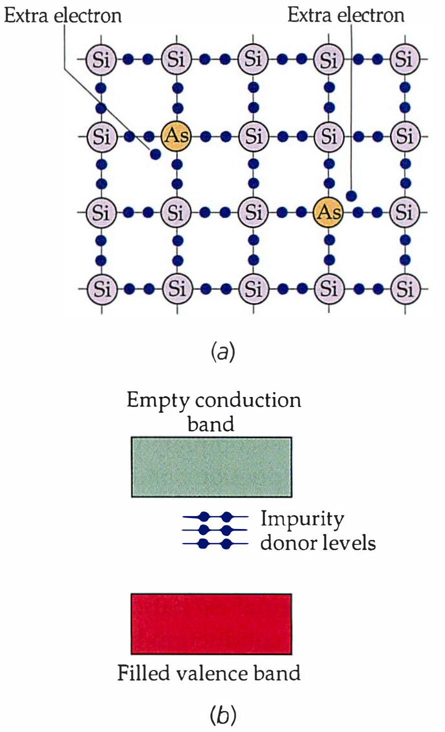


FIGURE 38-16 (*a*) A two-dimensional schematic illustration of silicon doped with arsenic. Because arsenic has five valence electrons, there is an extra, weakly bound electron that is easily excited to the conduction band, where it can contribute to electrical conduction. (*b*) Band structure of an *n*-type 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.

The effect on the band structure of a silicon crystal achieved by doping it with arsenic is shown in Figure 38-16*b*. 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 negative 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 38-17*a*). 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 38-17*b*. 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 these 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 positive holes. The fact that conduction is due to the motion of positive holes can be verified by the Hall effect. (The Hall effect is discussed in chapter 26.)

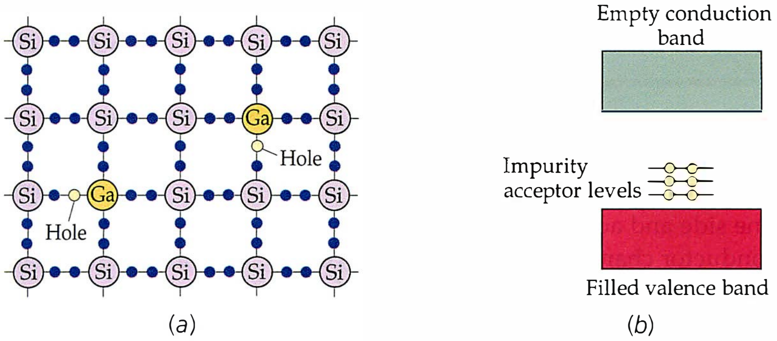
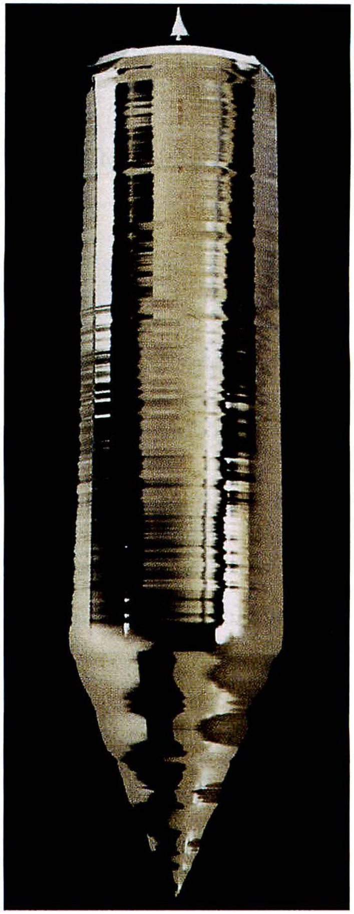


FIGURE 38-17 (*a*) A two-dimensional schematic illustration of silicon doped with gallium. Because gallium has only three valence electrons, there is a hole in one of its bonds. As electrons move into the hole the hole moves about, contributing to the conduction of electrical current. (*b*) Band structure 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.



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.

*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 **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 with its net positive charge will be at a higher potential than the p side with its 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.

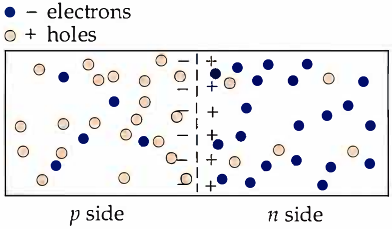


FIGURE 38-18 A pn junction. Because of the difference in their concentrations on either side of the pn junction, holes diffuse from the p side to the n side, and electrons diffuse from the n side to the p side. As a result, there is a double layer of charge at the junction, with the p side being negative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

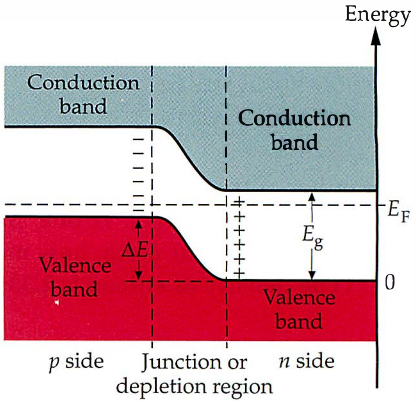


FIGURE 38-19 Electron energy levels for a pn junction.

*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

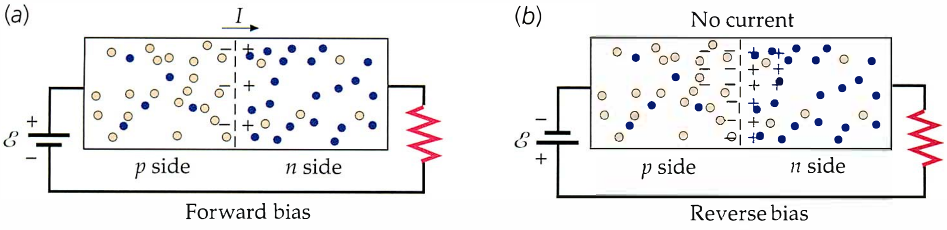


FIGURE 38-20 A *pn*-junction diode. (*a*) Forward-biased *pn* junction. The applied potential difference enhances the diffusion of holes from the *p* side to the *n* side and of electrons from the *n* side to the *p* side, resulting in a current *I*. (*b*) Reverse-biased *pn* junction. The applied potential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

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. A single-junction semiconductor device is called a **diode.**[†] Diodes have many uses. One 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.

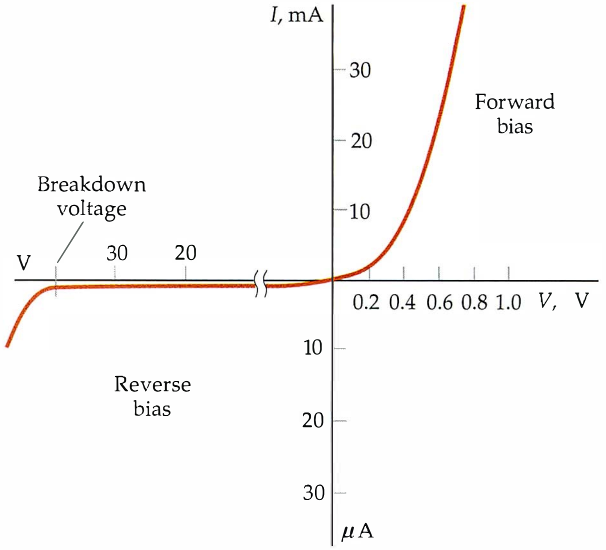
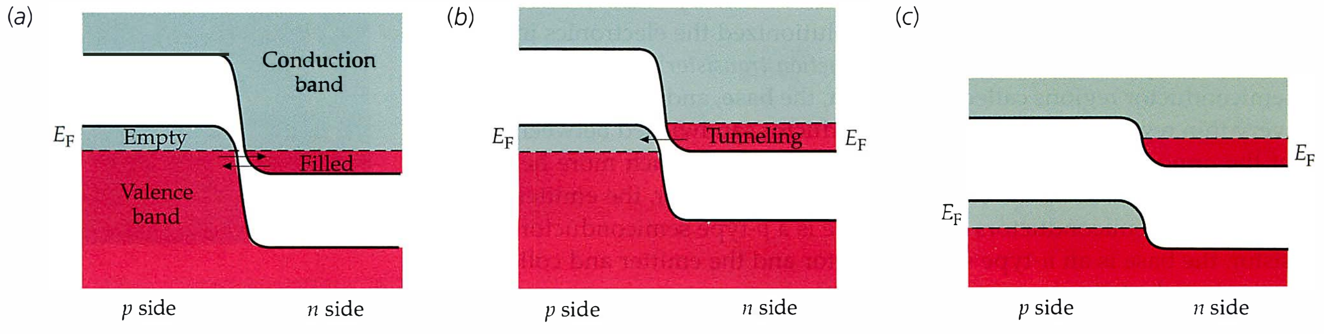


FIGURE 38-21 Plot of current versus applied voltage across a *pn* junction. Note the different scales on both axes for the forward and reverse bias conditions.

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 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 with 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 38-22*b*, and the tunneling of electrons from the

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.



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

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 38-22*c*, 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 curve versus the 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 curve versus the 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 number 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 with 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 with a large forward bias that produces a large excess concentration of electrons on the *p* side and holes on the *n* side of the junction. Under these conditions, the diode 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.

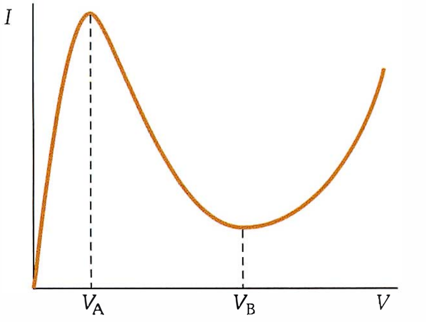


FIGURE 38-23 Current versus applied (bias) voltage V for a tunnel diode. For $V < V_A$, an increase in the bias voltage V enhances tunneling. For $V_A < V < V_B$, an increase in the bias voltage inhibits tunneling. For $V > V_B$, the tunneling is negligible, and the diode behaves like an ordinary *pn*-junction diode.

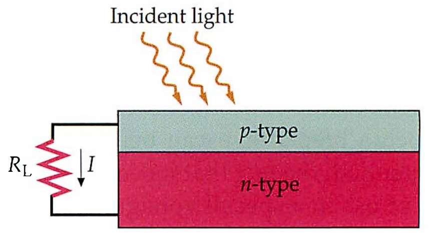


FIGURE 38-24 A *pn*-junction semiconductor as a solar cell. When light strikes the *p*-type region, electron–hole pairs are created, resulting in a current through the load resistance R_L.

38-8 Superconductivity

There are some materials 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 4.2 K) and put his technique to work exploring the properties of materials at temperatures in this 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 goes 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). To date (April 2003), 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 with 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.

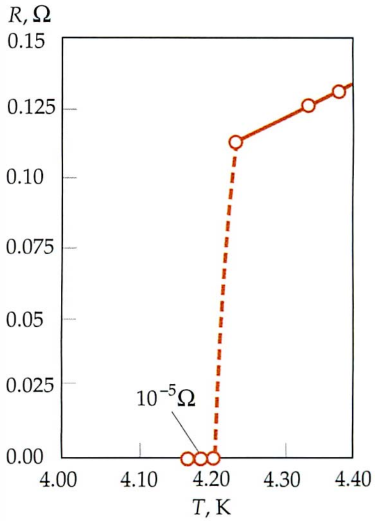


FIGURE 38-29 Plot by H. Kamerlingh Onnes of the resistance of mercury versus temperature, showing the sudden decrease at the critical temperature of $T = 4.2$ K.

The 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 superconduct at even higher temperatures.