from Elementary Modern Physics by Tipler

6-5 Band Theory of Solids

We have seen that, if the electron gas is treated as a Fermi gas and the electron–lattice collisions are treated as the scattering of electron waves, the free-electron model gives a good account of the thermal and electrical properties of conductors. However, this simple model gives no indication of why one material is a conductor and another is an insulator. Resistivities vary enormously between insulators and conductors. For example, the resistivity of a typical insulator, such as quartz, is of the order of $10^{16}\ \Omega\cdot\text{m}$, whereas that of a typical conductor is of the order of $10^{-8}\ \Omega\cdot\text{m}$. To understand why some materials conduct and others do not, we must refine the free-electron model and 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. As we have seen, the allowed energy levels in an isolated atom are often far apart. For example, in hydrogen, the energy for $n = 1$ is -13.6 eV and that for $n = 2$ is $(-13.6\text{ eV})/4 = -3.4$ eV. Let us consider two identical hydrogen atoms, and focus our attention on one particular energy level, such as $n = 2$. When the atoms are far apart, the energy of this level is the same for each atom. As the atoms are brought close together, the energy of this level for each atom changes because of the influence of the other atom. As a result, the $n = 2$ level splits into two levels of slightly different energies for the two-atom system.

If we have N identical atoms, a particular energy level in the isolated atom splits into N different, nearly equal energy levels when the atoms are close together. Figure 6-13 shows the energy splitting of the 1s and 2s energy levels for six atoms as a function of the separation of the atoms. In a macroscopic solid, N is very large—of the order of Avogadro's number—so each energy level splits into a very large number of levels called a **band.** Because the number of levels in the band is so large, 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.

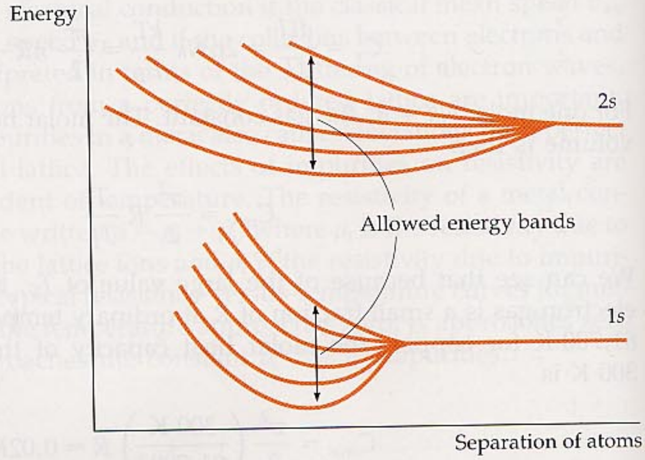


Figure 6-13 Energy splitting of the 1s and 2s energy levels for six atoms as a function of the separation of the atoms.

We can now understand why some solids are conductors and others are insulators. Consider sodium. There is room for two electrons in the 3s state of each atom, but each separate sodium atom has only one 3s electron. Therefore, when N sodium atoms are bound in a solid, the 3s energy band is only half filled. In addition, the empty 3p band overlaps the 3s band. The allowed energy bands of sodium are shown schematically in Figure 6-14. The occupied levels are shaded. We can see that many allowed energy states are available just above the filled ones, so the valence electrons can easily be raised to a higher-energy state by an electric field. Accordingly, sodium is a good conductor. Magnesium, on the other hand, has two 3s electrons, so the 3s band is filled. However, like sodium, the empty 3p band overlaps the 3s band, so magnesium is also a conductor.

The band structure of an ionic crystal, such as NaCl, is quite different. The energy bands arise from the energy levels of the Na^+ and Cl^- ions. Both of these ions have a closed-shell configuration, so the highest occupied band in NaCl is completely full. The next allowed band, which is empty, arises from the excited states of Na^+ and Cl^-. There is a large energy gap between the filled band and this empty band. A typical electric field applied to NaCl will be too weak to excite an electron from the upper energy levels of the filled band across the large gap into the lower energy levels of the empty band, so NaCl is an insulator. When an applied electric field is sufficiently strong to cause an electron to be excited to the empty band, the phenomenon is called dielectric breakdown.

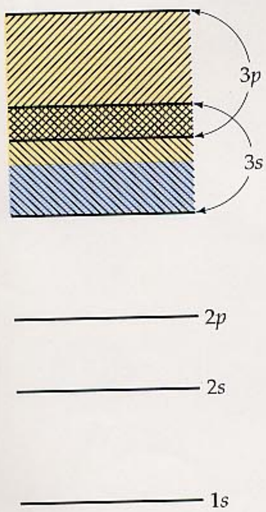


Figure 6-14 Energy band structure of sodium. The empty 3p band overlaps the half-filled 3s band. Just above the filled states are many empty states into which electrons can be excited by an electric field, so sodium is a conductor.

Figure 6-15 shows four possible kinds of band structures for a solid. The band occupied by the outermost, valence electrons is called the **valence band.** The lowest band in which there are unoccupied states is called the **conduction band.** In sodium, the valence band is only half filled, so the valence band is also the conduction band. In magnesium, the filled 3s band and the empty 3p bands overlap, forming a combined valence–conduction band that is only partially filled, so magnesium is also a conductor.

The band structure for a conductor such as copper is shown in Figure 6-15*a*. The lower bands are filled with the inner electrons of the atoms. According to the Pauli exclusion principle, no more electrons can occupy levels in these bands. The uppermost band that contains electrons is only about half full. Because these electrons are the electrons that form the metallic bond, this band is the valence band. In the normal state, at low temperatures, the lower half of the valence band is filled and the upper half is empty, so this band is also the conduction band. At higher temperatures, a few of the electrons are in the higher energy states of this band because of thermal excitation, but there are still many unfilled energy states above the filled ones. 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.

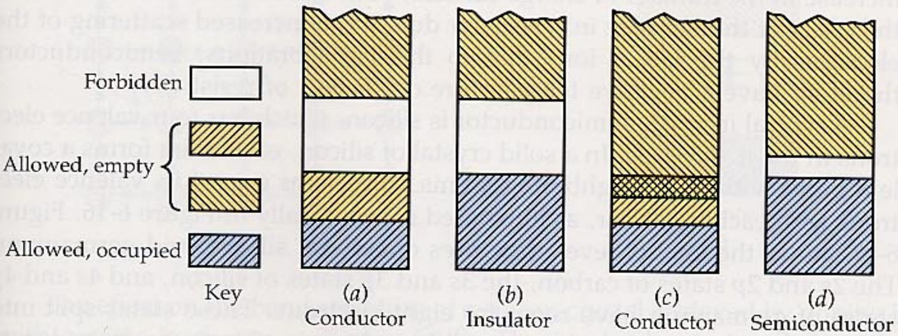


Figure 6-15 Four possible band structures for a solid. (*a*) A typical conductor. The valence band is only partially full, so electrons can be easily excited to nearby energy states. (*b*) A typical insulator. There is a forbidden band with a large energy gap between the filled valence band and the conduction band. (*c*) A conductor in which the allowed energy bands overlap. (*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.

Figure 6-15*b* shows the band structure for a typical insulator. At $T = 0$ K, the highest energy band that contains electrons is completely full. The next energy band containing empty energy states, the conduction band, is separated from the last filled band by an 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 because the energy gap is large compared with the energy an electron might obtain by thermal excitation, which on the average is of the order of $kT \approx 0.026$ eV at $T = 300$ K. This energy gap is sometimes referred to as the **forbidden energy band.** Very few electrons can be thermally excited to the nearly empty conduction band, even at fairly high temperatures. When an electric field 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.

In Figure 6-15*c* the valence and conduction bands overlap. A material such as magnesium with this type of band structure is a conductor. semi metal.

In some materials the energy gap between the top filled band and the empty conduction band is very small, as shown in Figure 6-15*d*. At $T = 0$, there are no electrons in the conduction band and the material is an insulator. 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.** 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 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 the motion of the electrons. The hole thus acts like a positive charge. An analogy of a two-lane, one-way road with one lane full of parked cars and the other empty may help to visualize the conduction of holes. 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 backwards 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 conductivity increases (and the resistivity decreases) as the temperature increases, which is contrary to the case for normal conductors. The reason is that as the temperature is increased, the number of free electrons is increased because there are more electrons in the conduction band. The number of holes in the valence band is also increased 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.

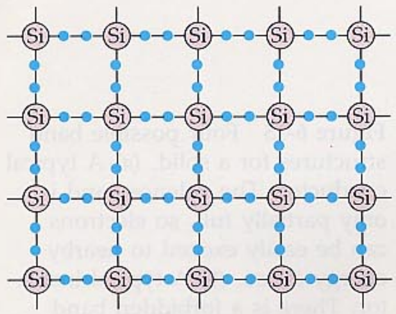


Figure 6-16 A two-dimensional schematic illustration of solid silicon. Each atom forms a covalent bond with four neighbors, sharing one of its four valence electrons with each neighbor.

A typical intrinsic semiconductor is silicon, which has four valence electrons in the $n = 3$ shell. In a solid crystal of silicon, each atom forms a covalent bond with four neighboring atoms and shares one of its valence electrons with each neighbor, as illustrated schematically in Figure 6-16. Figure 6-17 shows the energy level structures of carbon, silicon, and germanium. The 2*s* and 2*p* states of carbon, the 3*s* and 3*p* states of silicon, and 4*s* and 4*p* states of germanium have room for eight electrons. These states split into two hybrid states, each with room for four electrons. The lower hybrid state is filled with the valence electrons and the upper one is empty. At the separation of about 0.15 nm for carbon, the energy gap between these states is about 7 eV, so carbon is an insulator. For silicon and germanium, the separation is about 0.24 nm and the energy gap is about 1 eV, so these elements are intrinsic semiconductors.

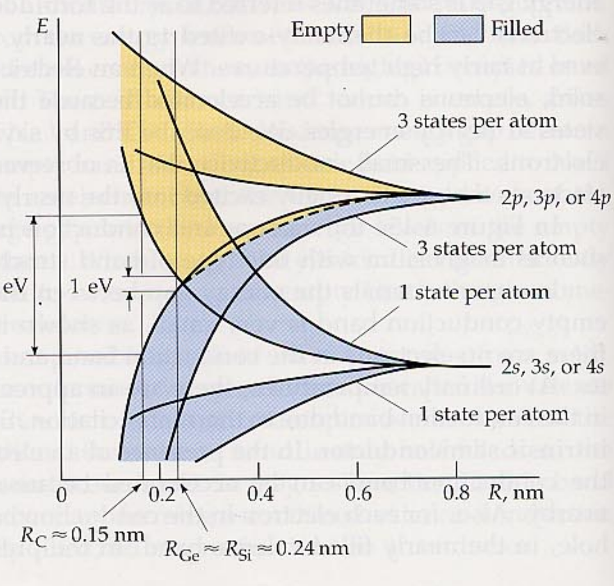
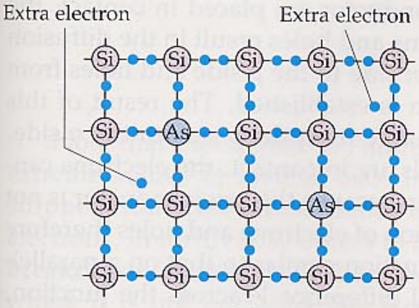


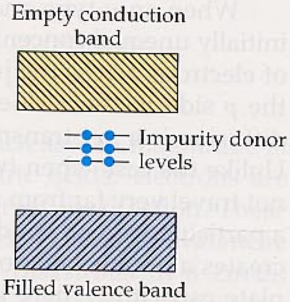
Figure 6-17 Splitting of the 2*s* and 2*p* states of carbon, the 3*s* and 3*p* states of silicon, or the 4*s* and 4*p* states of germanium versus the separation of the atoms. The energy gap between the filled states and empty states is about 7 eV for carbon but only about 1 eV for silicon and germanium.

6-6 Impurity Semiconductors

Most semiconductor devices, 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 6-18*a* is a schematic illustration of silicon doped with a small amount of arsenic such that arsenic atoms replace a few of the silicon atoms in the crystal lattice. Arsenic has five electrons in its valence shell rather than the four of silicon. Four of these electrons take part in covalent 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.



(*a*)

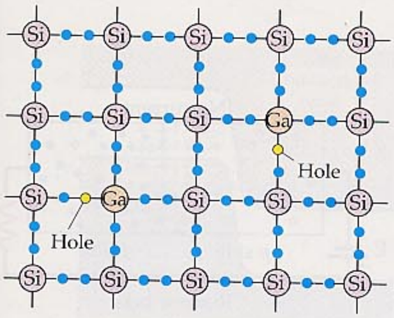


(*b*)

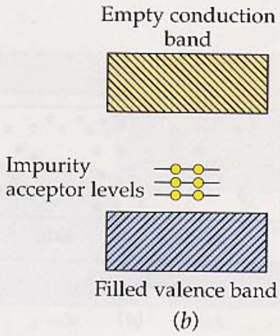
Figure 6-18 (*a*) A two-dimensional schematic illustration of silicon doped with arsenic. Because arsenic has five valence electrons, there is an extra, weakly bound electron that is easily excited to the conduction band, where it can contribute to electrical conduction. (*b*) Band structure of an *n*-type 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 6-18*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 that has 3 electrons in its valence level (Figure 6-19*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 6-19*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.



(*a*)



(*b*)

Figure 6-19 (*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.

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 holes can be verified by the Hall effect.

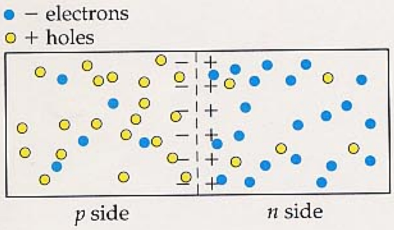


Figure 6-20 A *pn* junction. Because of the difference in their concentrations, holes diffuse from the *p* side to the *n* side and electrons diffuse from the *n* side to the *p* side. As a result, there is a double layer of charge at the junction with the *p* side being negative and the *n* side being positive.

6-7 Semiconductor Junctions and Devices

Semiconductor devices such as diodes and transistors make use of *n*-type and *p*-type semiconductors joined together as shown in Figure 6-20. In practice, the two types of semiconductors are often a single silicon crystal doped with donor impurities on one side and acceptor impurities on the other. The region in which the semiconductor changes from a *p*-type to an *n*-type is called a **junction.**

When an *n*-type 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, there will be very few charge carriers of either type, so the junction region has a high resistance. Figure 6-21 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.

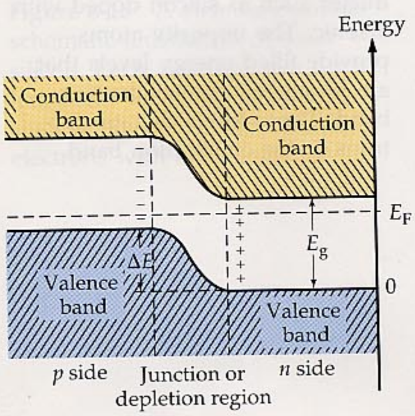


Figure 6-21 Electron energy levels for an unbiased *pn* junction.

A semiconductor with a *pn* junction can be used as a simple diode rectifier. In Figure 6-22, an external potential difference has been applied across the junction by connecting a battery and resistor to the semiconductor. When the positive terminal of the battery is connected to the *p* side of the junction, as shown in Figure 6-22*a*, the diode 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 6-22*b*, the diode is said to be **reverse biased.** Reverse biasing tends to increase the potential difference across the junction, thereby further inhibiting diffusion. Figure 6-23 shows a plot of current versus voltage for a typical semiconductor junction. Essentially, the junction conducts only in one direction, the same as a vacuum-tube diode. Junction diodes have replaced vacuum diodes in nearly all applications except when a very high current is required.

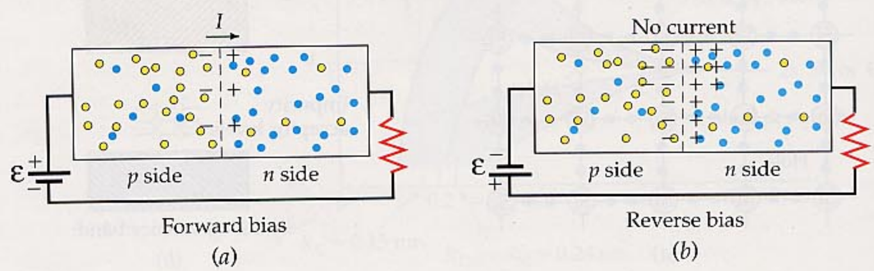


Figure 6-22 A *pn*-junction diode. (*a*) Forward-biased *pn* junction. The applied potential difference enhances the diffusion of holes from the *p* side to the *n* side and electrons from the *n* side to the *p* side, resulting in a current I. (*b*) Reverse-biased *pn* junction. The applied potential difference inhibits the further diffusion of holes and electrons, so there is no current.

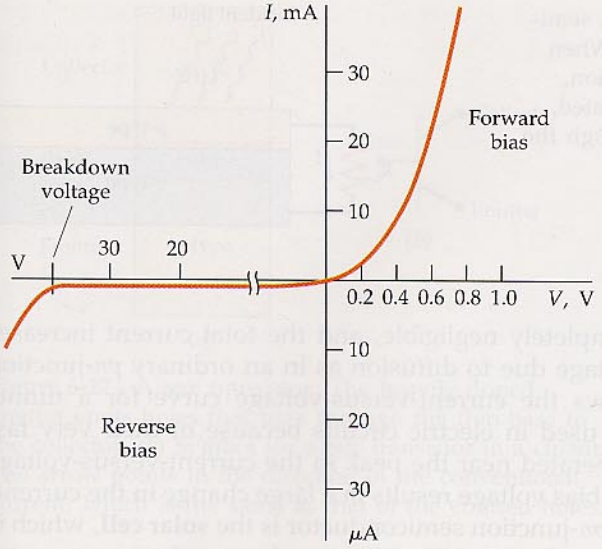


Figure 6-23 Current versus applied voltage across a *pn* junction. Note the different scales for the forward and reverse bias conditions.

Note that the current in Figure 6-23 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 sharp voltage value makes it of use in a special voltage reference standard known as a **Zener diode.**

An interesting effect that we can discuss only qualitatively occurs if both the *n* side and *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 6-24*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. This flow of electrons 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 6-24*b*, and the tunneling of electrons from the *n* side to the *p* side is increased whereas that 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 6-24*c*, and the tunneling current is decreased. Although the diffusion current is increased, the net current is decreased. At large bias voltages, the

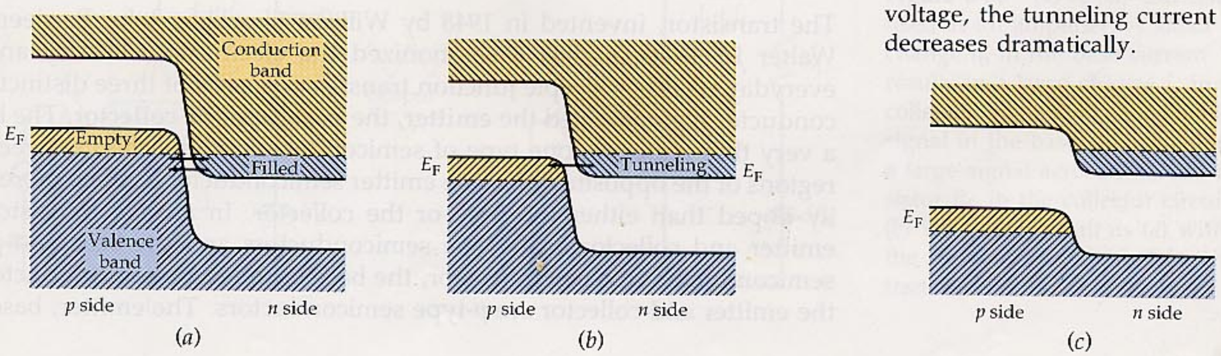


Figure 6-24 Electron energy levels for a heavily doped *pn*-junction tunnel diode. (*a*) With no bias voltage, some electrons tunnel in each direction. (*b*) With a small bias voltage, the tunneling current is enhanced in one direction, making a sizable contribution to the net current. (*c*) With further increases in the bias voltage, the tunneling current decreases dramatically.

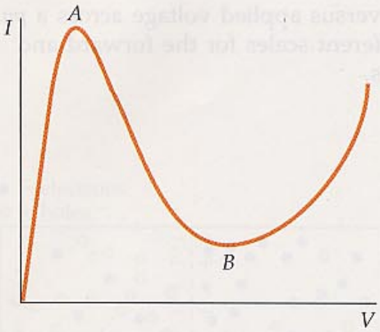
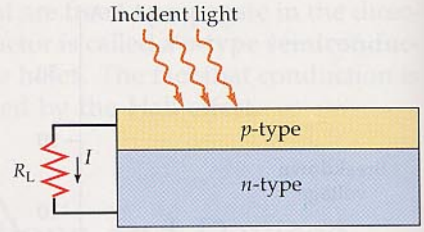


Figure 6-25 Current versus applied voltage for a tunnel diode. Up to point *A*, an increase in the bias voltage enhances tunneling. Between points *A* and *B*, an increase in the bias voltage inhibits tunneling. After point *B*, the tunneling is negligible, and the diode behaves like an ordinary *pn*-junction diode.

Figure 6-26 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.



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 6-25 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 6-26. 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 excess positive charge in the *p*-type region. The result is a potential difference between the two regions, which in practice is about 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.

A light-emitting diode (LED).

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 displays for digital watches and calculators.

Transistors

The transistor, invented in 1948 by William Shockley, John Bardeen, and Walter H. Brattain, has revolutionized the electronics industry and our everyday world. A simple 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 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. The emitter, base, and

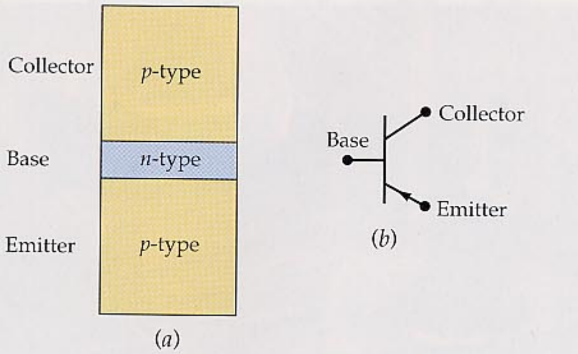


Figure 6-27 A *pnp* transistor. The heavily doped emitter emits holes that pass through the thin base to the collector. (*b*) Symbol for a *pnp* transistor in a circuit. The arrow points in the direction of the conventional current, which is the same as that of the emitted holes.

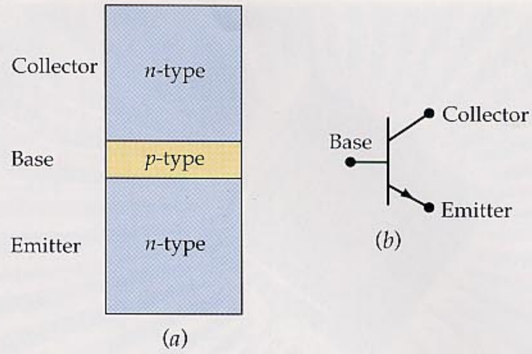


Figure 6-28 An *npn* transistor. The heavily doped emitter emits electrons that pass through the thin base to the collector. (*b*) Symbol for an *npn* transistor. The arrow points in the direction of the conventional current, which is opposite the direction of the emitted electrons.

collector behave somewhat similarly to the cathode, grid, and plate in a vacuum-tube triode, except that in a *pnp* transistor it is holes that are emitted rather than electrons.

Figures 6-27 and 6-28 show, respectively, a *pnp* transistor and an *npn* transistor with the symbols used to represent each transistor in circuit diagrams. We see that a transistor consists of two *pn* junctions. We will discuss the operation of a *pnp* transistor. The operation of an *npn* transistor is similar.

In normal operation, the emitter–base junction is forward biased, and the base–collector junction is reverse biased, as shown in Figure 6-29. The heavily doped *p*-type emitter emits holes that flow across the emitter–base junction into the base. Because the base is very thin, most of these holes flow across the base into the collector. This flow constitutes a current I_c from the emitter to the collector. However, some of the holes recombine in the base producing a positive charge that inhibits the further flow of current. 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 circuit connected to the base. In Figure 6-29, 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

$$I_c = \beta I_b \tag{6-35}$$

where β is called the **current gain** of the transistor. Transistors can be designed to have values of β as low as 10 or as high as several hundred.

Figure 6-30 shows a simple *pnp* transistor used as an amplifier. A small time-varying input voltage v_s is connected in series with a bias voltage V_{eb}.

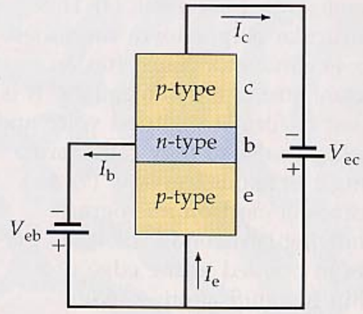


Figure 6-29 A *pnp* transistor biased for normal operation. Holes from the emitter can easily diffuse across the base, which is only tens of nanometers thick. Most of the holes flow to the collector, producing the current I_c.

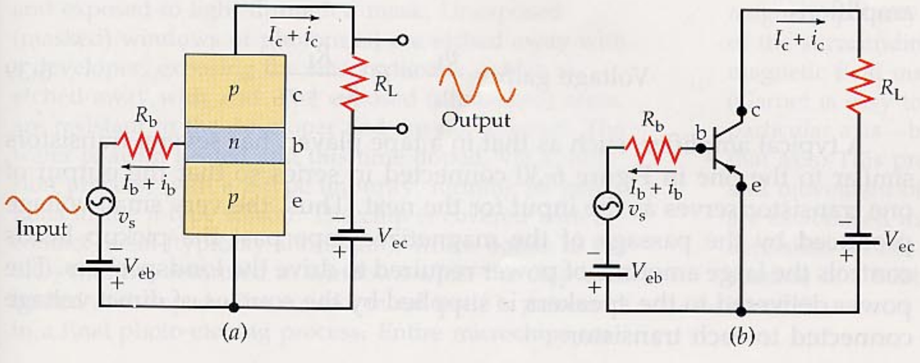
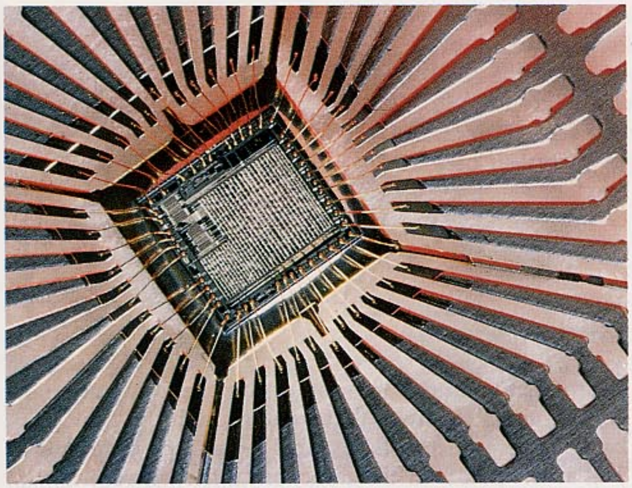
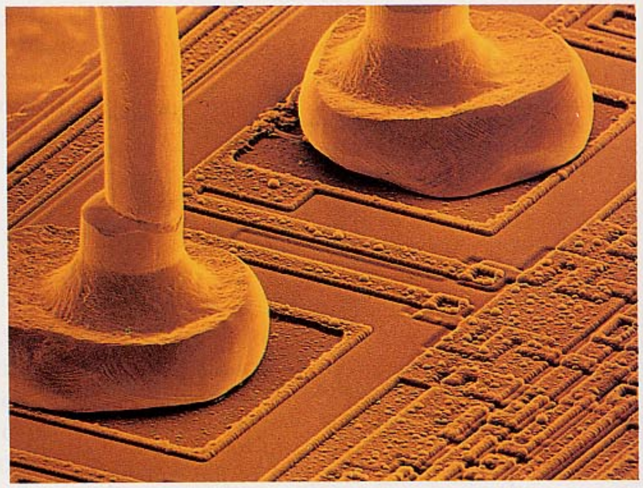


Figure 6-30 (*a*) A *pnp* transistor used as an amplifier. A small change i_b in the base current results in a large change i_c in the collector current. Thus a small signal in the base circuit results in a large signal across the load resistor R_L in the collector circuit. (*b*) The same circuit as (*a*) with the conventional symbol for the transistor.



(a)



(b)

Integrated circuits (often called ICs or chips) combine "active" electronic devices (transistors and diodes) with "passive" ones (capacitors and resistors) on a single semiconductor crystal. (a) This particular chip, shown connected to 44 conductor leads, has an actual size of 6.4 mm square. It is used to format digitized voice and data signals, so they can share a single transmission line. (b) A scanning electron micrograph showing two conductor leads precision bonded to the edge of a chip (magnification: ×163).

The base current is then the sum of a steady current I_b produced by the bias voltage V_{eb} and a 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 direct current $I_c = \beta I_b$ and an alternating current $i_c = \beta i_b$. We thus have a current amplifier in which the time-varying output current i_c is β times 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:

$$i_b = \frac{v_s}{R_b + r_b} \qquad 6\text{-}36$$

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

$$v_L = i_c R_L \qquad 6\text{-}37$$

Using Equation 6-35, we have

$$i_c = \beta i_b = \beta \frac{v_s}{R_b + r_b}$$

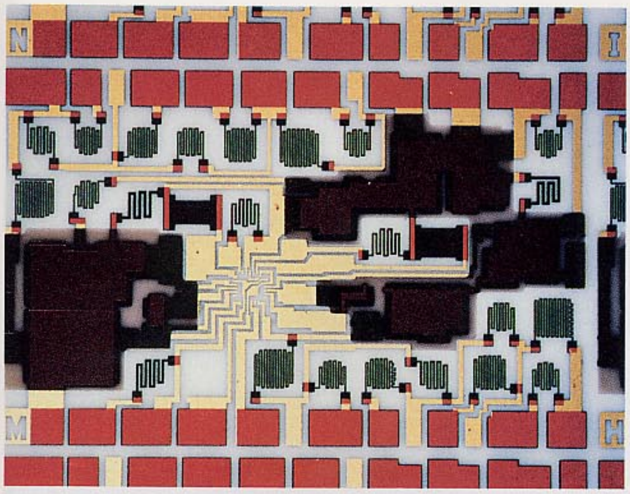
The output voltage is thus related to the input voltage by

$$v_L = \beta \frac{R_L}{R_b + r_b} v_s \qquad 6\text{-}38$$

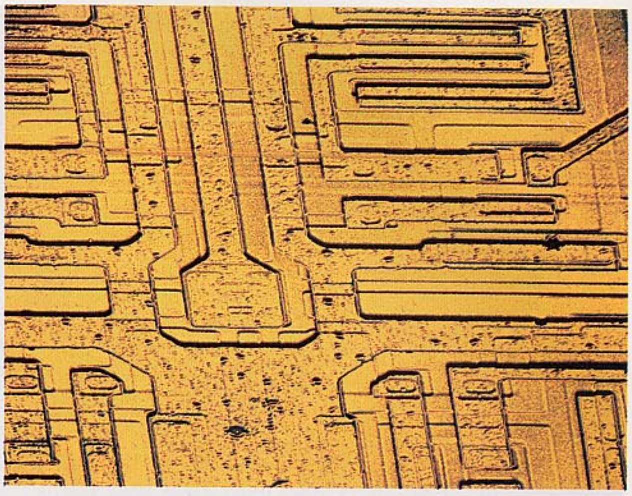
The ratio of the output voltage to the input voltage is the **voltage gain** of the amplifier:

$$\text{Voltage gain} = \frac{v_L}{v_s} = \beta \frac{R_L}{R_b + r_b} \qquad 6\text{-}39$$

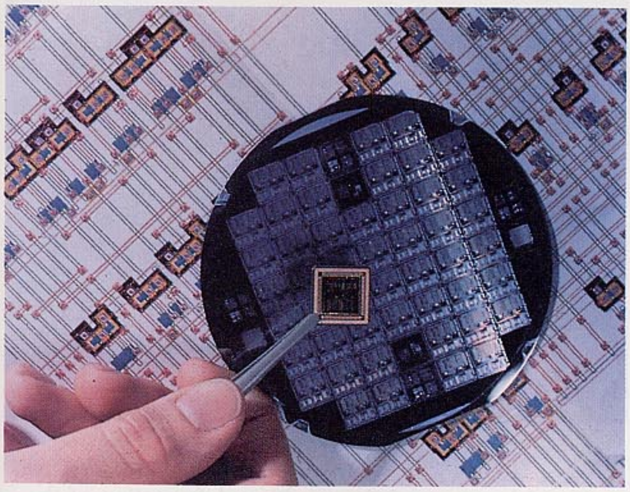
A typical amplifier, such as that in a tape player, has several transistors similar to the one in Figure 6-30 connected in series so that the output of one transistor serves as the input for the next. Thus, the very small voltage produced by the passage of the magnetized tape past the pickup heads controls the large amounts of power required to drive the loudspeakers. The power delivered to the speakers is supplied by the sources of direct voltage connected to each transistor.



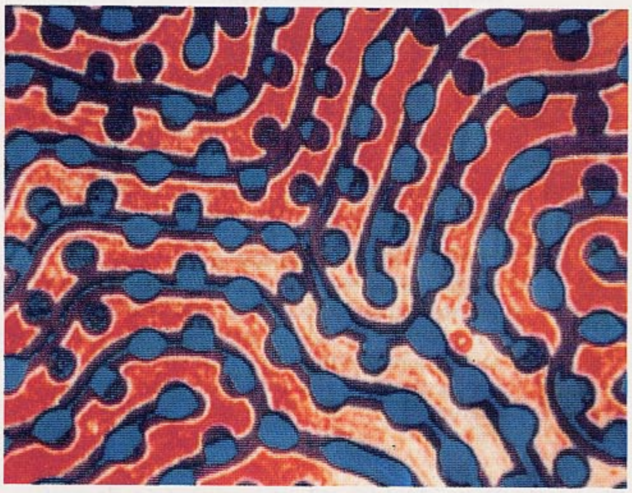
(c)



(d)

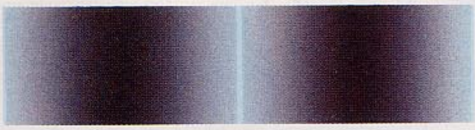


(e)

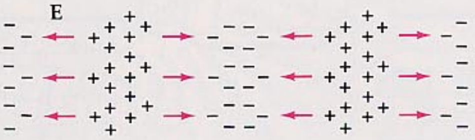


(f)

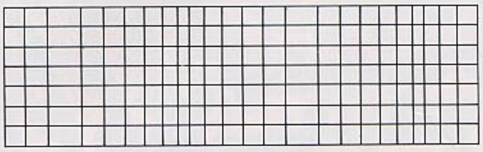
(c) Capacitors (orange blocks), resistors (brown blocks and meandering black lines), and conductors (gold lines) on a ceramic base, formed here by metal films only a few tenths of a micrometer thick. No means have been found to directly fabricate inductors (the other passive circuit component) on ICs; they are simulated with other circuitry or appended to a chip as discrete components. (d) A scanning electron micrograph of metal oxide silicon (MOS) transistors in patterned layers (magnification ×106). MOS transistors are manufactured by heating an original silicon wafer to about 1000°C, causing a layer of silicon dioxide (SiO_2) to form on its surface. This is coated with a photoresist and exposed to light through a mask. Unexposed (masked) windows of photoresist are etched away with a developer, exposing the silicon dioxide, which is etched away with acid. The exposed (unmasked) areas are resistant to the developer and are not affected. The wafer is again heated and this time doped, via a diffusion process, with a *p*-type impurity, forming *pn* junctions in the *n*-type silicon. The chip is covered with a contact metal (typically aluminum), which bonds to the SiO_2 that has reformed in windows while the chip was heated and doped. The contact metal itself is patterned in a final photo-etching process. Entire microchips are fabricated by an elaboration, using many masks, of this process. (e) The chip in the tweezers holds 150,000 transistors. Beneath it is a 4-inch wide silicon wafer, awaiting dicing, on which a group of chips have been fabricated simultaneously. In the background is a detail of the "stare plot," the layout of the chip's circuits. (f) Magnetized domains ("bubbles"), blue in this video micrograph, flow along channels in a thin-film garnet memory crystal. Magnetic bubble memory chips are the integrated-circuit analog to magnetic recording tape and disks. The bubbles (actually, cylinders seen in cross section) are created when the garnet is placed between two permanent magnets. They represent regions whose magnetic polarity points in a direction opposite to that of the surrounding crystal. An additional external magnetic field manipulates the position of the bubbles. (Garnet is easy to magnetize, up or down, along a particular axis—but hard to magnetize perpendicular to that axis. This property is necessary for the formation and movement of bubbles.) Storage sites for bubbles are established using a layer of ferromagnetic material deposited on the surface of the crystal; the presence or absence of a bubble at a site can be used to represent a bit of data.



A fixed pattern of light and dark is established in a photorefractive crystal, for instance by letting two laser beams of the same frequency interfere within the crystal.

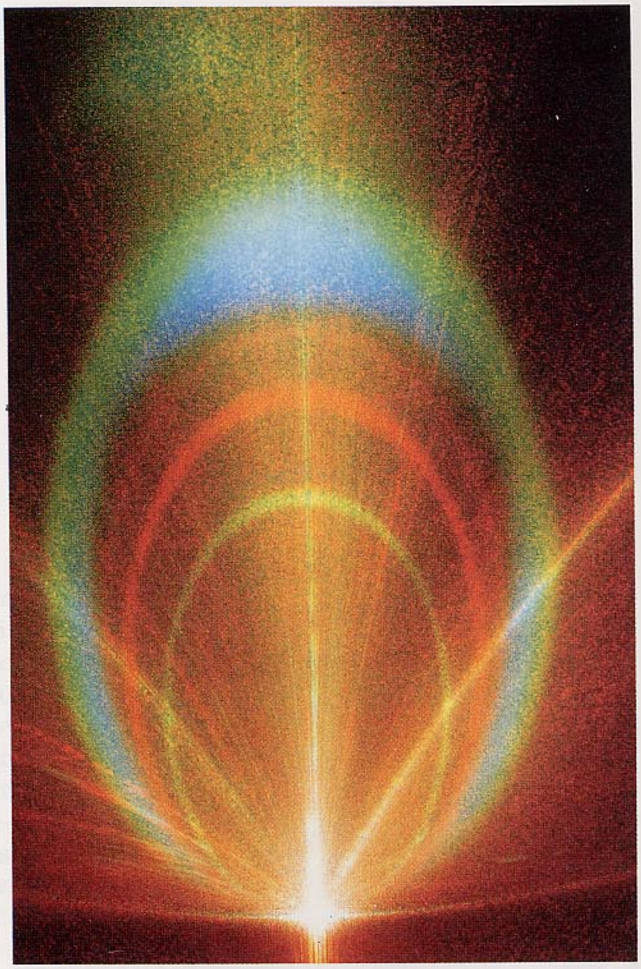


Responding to the electric fields of the light, free electrons or holes move away from bright regions of the crystal. The resulting build-up of charge creates a static electric field within the crystal.



The static electric field acts on the atoms of the crystal lattice. The positions of the atoms shift and the lattice becomes distorted. Distortions in the lattice cause light to propagate faster in some regions of the crystal and slower in others: that is, the refractive index of the crystal is altered.

(*a*)



(*b*)

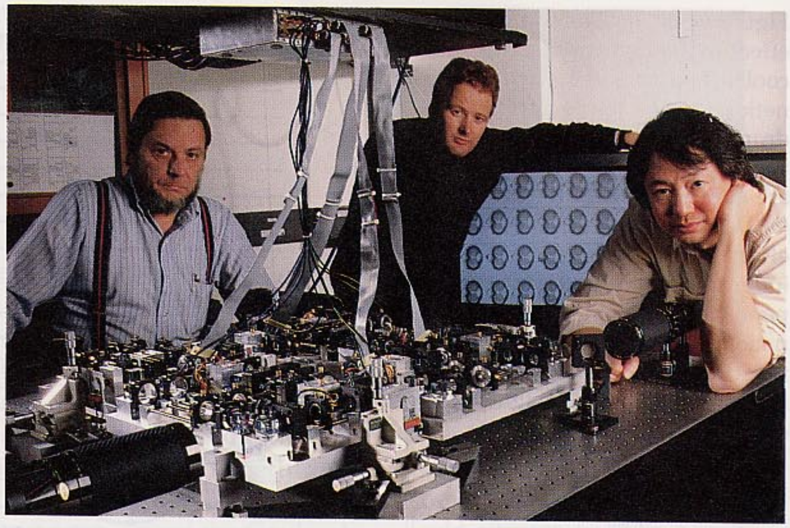
Photorefractivity. Light incident on a certain class of materials called "photorefractive crystals" causes a rearrangement of their crystal lattice, which in turn alters the optical properties of that lattice. This behavior is due to impurities or defects in the photorefractive crystal. These defects are a source of free electrons or holes that migrate away from illuminated areas in the crystal. The resulting rearrangement of free charge gives rise to electric fields that act on the atoms in the crystal lattice, distorting its shape. This process is summarized in schematic (*a*). Here, the incident light is from two interfering laser beams. The interference pattern of these beams is translated into a periodic pattern in the crystal lattice itself. Such a pattern is sometimes referred to as a "refractive-index grating." Typically, the grating pattern will be phase shifted with respect to the interference pattern of the light beams. One of the beams will tend to interfere constructively with light scattered by the grating, the other destructively. As a result, the first beam will be intensified passing through the crystal, and the second beam will be reduced. Figure (*b*) shows the result of the "beam-fanning effect" occurring in the photorefractive crystal barium titanate ($BaTiO_3$). Multicolored incident laser beams are partially scattered by defects in the crystal. An interference pattern arises between the incident and scattered beams, and the resulting deformation of the crystal lattice changes the lattice's index of refraction. The incident beams are refracted to a different angle and a new interference pattern arises, causing a further change in the index of refraction.

Beam fanning lies at the heart of new devices called "phase-conjugate mirrors." Such mirrors produce beams that retrace the path of any incident beam exactly, with the leading edge of the incident wavefront appearing last in the phase-conjugate beam. (In reflection from an ordinary mirror the leading edge of the incident wavefront emerges first.) Incident light undergoing beam fanning in a photorefractive material may be preferentially swept into a corner of the crystal where it is twice internally reflected. The path of such light forms a loop, in which the exiting beam is phase conjugate to the incoming beam. Such a configuration is shown in (*c*) inside a crystal of barium titanate. The volume refractive-index grating generated in the crystal by these loops produces the phase-conjugate beam. A phase-conjugate beam can be used to restore its incident beam if the incident beam has been overlaid with noise, or in other ways distorted.

Mutually reinforcing beams can give rise to spectacular effects. For instance, if a laser beam is sent

Continued

(c)



(d)

separately into each of two crystals positioned, say, a meter apart, a third beam of light will, after a few seconds, spontaneously arise between the two crystals—connecting them. Light scattered by each crystal is phase conjugated by the other crystal. Such a phase-conjugate beam is aimed directly at the other crystal and so contributes to a growing beam connecting the two.

Refractive index gratings, if formed from a reference beam interfering with a beam scattered from an object, are effectively holograms stored in the photorefractive material. Such holograms are often referred to as being "real-time," since they can be generated and erased continuously in a crystal. The use of applied light to control the transmission of light in crystals with impurities is generally reminiscent of the use of applied voltages to control electrical conduction and electrical properties of semiconductors. The hope thus arises that an "integrated optics" technology can be developed that might supersede the microelectronics of integrated circuits. Early prototypes along these lines have, in fact, been developed—one of which is shown in (d).

6-8 Superconductivity

There are some materials for which the resistivity is zero below a certain temperature, called the **critical temperature** T_c. This phenomenon, called **superconductivity,** was discovered in 1911 by the Dutch physicist H. Kamerlingh Onnes. Figure 6-31 shows his plot of the resistance of mercury versus temperature. The critical temperature for mercury is 4.2 K. The critical temperature varies from material to material, but below this temperature the resistance of the material is zero. Critical temperatures for other superconducting elements range from less than 0.1 K for hafnium and iridium to 9.2 K for niobium. The critical temperatures of various superconducting materials are given in Table 6-3 on page 227. In the presence of a magnetic field B, the critical temperature is lower than it is when there is no field. As the magnetic field increases, the critical temperature decreases. If the magnetic field is greater than some critical field B_c, superconductivity does not exist at any temperature.

Many metallic compounds are also superconductors. For example the superconducting alloy Nb_3Ge, discovered in 1973, has a critical temperature of 23.2 K, which was the highest known until 1986. Despite the cost and inconvenience of refrigeration with expensive liquid helium, which boils at 4.2 K, many superconducting magnets were built using such materials, because such magnets produce no heat.

The conductivity of a superconductor cannot be defined since its resistance is zero. There can be a current in a superconductor even when the electric field in the superconductor is zero. Indeed, steady currents have

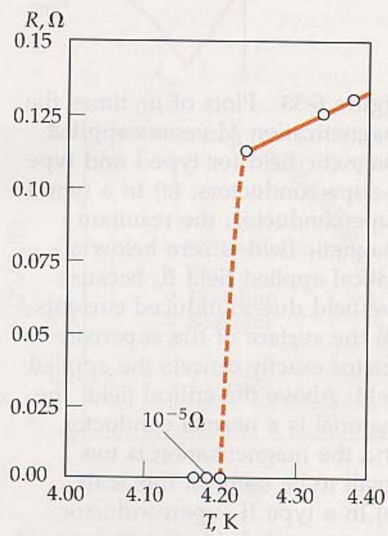


Figure 6-31 Plot by Kamerlingh Onnes of the resistance of mercury versus temperature, showing sudden decrease at the critical temperature $T = 4.2$ K.