CHAPTER 16

Solids Matter

16.1 THE SUCCESS OF QUANTUM MECHANICS

The triumph of quantum mechanics during the 1920s opened the door to a new understanding of a wide range of phenomena involving the most fundamental properties of matter and radiation, the fundamental constituents of the Universe. The new physics also enabled an amazing variety of technological applications that are continuing to transform our lives—from lasers and microchip transistors to new medical imaging devices, and a host of new materials. Combined with relativity theory, quantum mechanics is the foundation for understanding the interactions of radiation and matter, the behavior of elementary particles at extremely high energies, and many of the stellar phenomena observed in cosmology. Soon after the formulation of quantum mechanics, scientists successfully applied the new theory to a new understanding of the nucleus of the atom that is still used today. This is discussed in Chapters 17 and 18. During the same period, scientists also utilized quantum mechanics to gain a new understanding of the structure and behavior of matter, including, in particular, the thermal and electrical properties of solids. These are the focus of the current chapter.

16.2 FORMING A SOLID

It was well known that matter appears in one of three states, or *phases*—solid, liquid, and gas or vapor. According to the kinetic-molecular theory (Chapter 7), one factor distinguishing each of these states is the amount of energy distributed among the molecules making up the material. In the gaseous state, the molecules have the most energy and in the solid state the least. Transitions from one state to another can be achieved by adding or extracting energy in the form of heat, as is readily observed. But that can't be the whole story; as we shall examine below, at each transition from one state to another, the heat content seems to disappear (if going to states of higher temperature) or to appear without a loss of energy elsewhere (if going to states of lower temperature)! The temperature of the material does not change at all. We need to include an additional concept, the concept of *entropy*, which gives an idea of the level of disorder. Much of the basic understanding of phase transitions and the formation of solids was developed around 1900, but a complete understanding required quantum mechanics over the subsequent decades, and some aspects of these complicated processes are still under investigation.

In order to view the basic process leading to the formation of a solid, we look at the behavior of water, which can also easily be observed in your laboratory. We start with liquid water, which we will cool down by extracting heat. As the water cools, the amount of decline in its temperature, ΔT, is related to the amount of heat extracted, ΔQ, by the expression (Chapter 6):

$$\Delta Q = mc\,\Delta T,$$

where m is the mass of the water and c is the specific heat of liquid water, which in this case is 1 calorie of heat per gram and per degrees centigrade.

As we continue to cool the water, we reach a temperature at which the liquid undergoes a *phase change* from liquid into solid. All liquids eventually freeze into solids, but they do so at very different temperatures. Water at atmospheric temperature will begin to freeze at 0°C (273 K), but liquid helium does not freeze until it reaches 4 K, very near to absolute zero. At this point we observe the effect noted earlier—as we extract heat, the liquid continues to freeze into ice, but the temperature stays the same. The heat extracted seems to have been "hidden" in the liquid and to be associated with the formation of the solid, ice. This "hidden" heat is called the *latent heat of fusion*, named for the linking of the molecules together into a

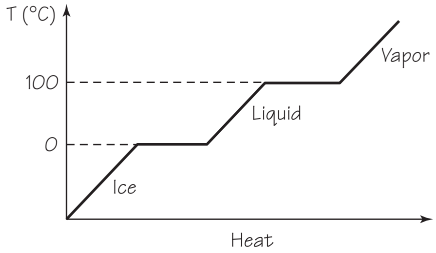


FIGURE 16.1 Three phases of water and its transitions as heat is added.

solid. (This has nothing to do with nuclear fusion.) For water, it is about 80 calories of heat per gram of water.

What is this latent heat and what does it have to do with the freezing of a liquid into a solid? Since the cooling process involves the extraction of energy from the material, the molecules begin to arrange themselves into a state of minimum energy. This is achieved as the molecules group together into an arrangement that is as orderly and as structured as possible. From the earlier discussion about entropy, you know that the state of minimum energy will have the highest order, which means that it has the lowest entropy. The change in the entropy, ΔS, is defined as the change in the heat content per unit of absolute temperature (see Chapter 6):

$$\Delta S = \frac{\Delta Q}{T}.$$

So, the latent heat that is extracted from the water as it freezes, ΔQ, is directly associated with the decreasing entropy of the water molecules as they form the ordered structure of solid ice. Of course, if heat is pumped back into the system, the entropy will start to increase as the molecules begin to vibrate faster and faster. They may eventually break their solid bonds. This is observed as melting. Again the temperature remains constant as the solid is melting under the input of heat energy. During this process, the incoming heat energy becomes latent once again, since the temperature does not change, but the entropy increases and the orderly arrangement within the solid breaks down, forming a pool of water where once we had an orderly ice cube.

As a solid is formed from a liquid, the type of orderly arrangement that it creates depends upon the types of molecules and the nature of the electrical attraction between the molecules. The lowest entropy states would involve the largest number of bonds possible between the molecules. This

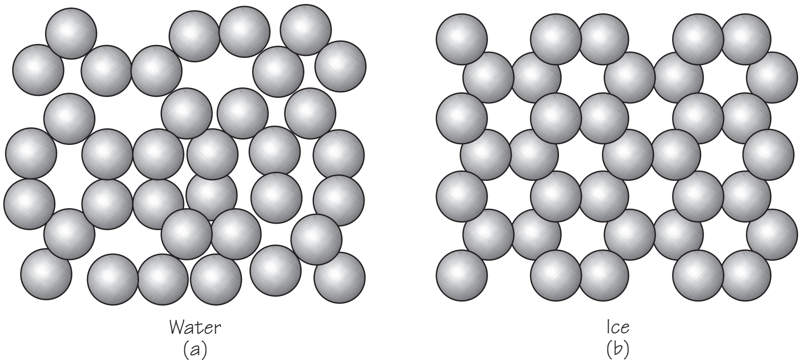


FIGURE 16.2 Schematic drawing of water molecules as liquid and solid.

occurs in the formation of crystals, which involve beautifully ordered arrays of billions and billions of molecules. In these arrays a basic crystal structure involving only a small number of molecules is repeated many times over, like grouted tiles on a floor, or a wallpaper pattern. Water, salt, and many metals form such crystalline structures. Many metal alloys, which involve a mixture of different metals, form what are called "polycrystalline" ("many crystal") solids. Examples include metal door knobs, tableware, and automotive parts.

However, not all solids involve crystalline bonding. There are three other important types of bonding (as well as some exotic types). These include *quasi-crystalline*, *amorphous*, and *composite* bonding. The types of molecules involved and even the way in which the solid was formed determine the type of structure formed. If the cooling from liquid occurs too quickly, the molecules which might normally form a perfect crystal do not have time to arrange themselves into a low-entropy crystal structure. They literally tend to be frozen into place while in the liquid state. In this case there is only local, short-range order and no long-range repetitive array. This structure is called *amorphous*. Some solids are naturally amorphous, even when frozen slowly. These include rubber, glass, and plastics. Carbon in the form of coal is amorphous, but in the form of graphite or diamond it is highly crystalline.

Quasi-crystalline solids were first discovered only in 1983. The molecules are arranged into a very orderly three-dimensional pattern, but the pattern does not repeat itself throughout the crystal as it does in a truly crystalline structure.

Finally, there are the *composite solids*, which involve the bonding together of different materials with different structures. Examples here include wood, cement, and fiberglass, as well as such biological materials as bone, flesh, and muscle. Research in recent years to develop materials that can act as biological "spare parts" has undergone a revolution because of the discovery of composite solids.

16.3 QUANTUM SOLIDS

One of the first recognized successes of the early quantum theory arose, not from the study of radiation, but from the theory of solids. Once again Albert Einstein was at the forefront. After presenting the hypothesis of light quanta in 1905, which explained the photoelectric effect (Section 13.5), Einstein showed in 1906 that Planck's earlier results on the distribution of energy in electromagnetic radiation could be derived using two assumptions. The first, already familiar from his 1905 work, was that light is composed of light quanta; the second was that the light quanta are emitted and absorbed by oscillating charges within matter that are vibrating only with certain quantized energies. This means that light is emitted by oscillating charges within matter, as might be expected from Maxwell's theory, but the energies both of the emitted light and of the oscillators are restricted to certain quantized values, which contradicted classical theory. The oscillators behaved like charges attached to springs, oscillating back and forth with certain frequencies corresponding to certain allowed quantum energies.

In 1904, the German physicist Paul Drude had proposed a model of a crystalline solid in which the atoms are arranged in a lattice such that each atom is bound to its neighboring atoms by bonds that can act like little oscillator springs. Drude showed, using classical physics, that the three-dimensional oscillation of the lattice atoms can account for the observed absorption of infrared radiation by some solids, if the atoms are assumed to be ionized. A similar oscillation of the electrons, attached to the atoms or ions, accounted for the observed absorption of ultraviolet radiation in solids.

In 1907 Einstein applied the quantum hypothesis to Drude's classical oscillators. Even before Drude's model, Einstein had been wondering as early as 1901 whether the internal kinetic energy of solids and liquids might be regarded as the energy of "electric resonators." One outstanding puzzle at that time involved the behavior of the specific heat, C, of solids as the temperature decreased. Normally the specific heat is assumed to be constant in the relationship between the transfer of heat and the change in temperature of a substance

$$\Delta Q = mc\,\Delta T.$$

Applying classical physics to Drude's oscillator model of the atoms in a lattice confirmed that the specific heat C should be a constant for solids at

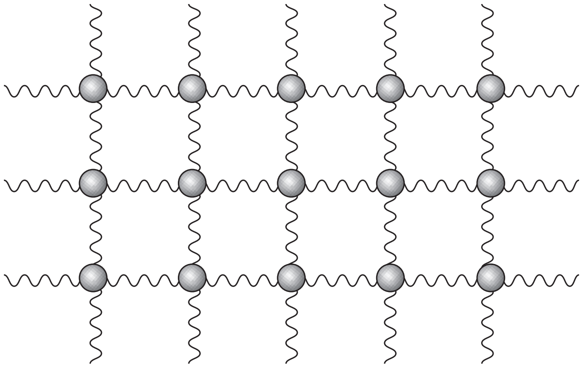


FIGURE 16.3 Drude's model of a crystalline solid.

all temperatures. However, when Einstein applied the quantum hypothesis to the oscillators in Drude's model, he obtained an expression for the specific heat indicating that the value of the specific heat should change with temperature. While remaining constant at room temperature and slightly below, the specific heat would drop precipitously to zero at the temperature declined to zero, as shown in Figure 16.3. Conversely, as the temperature increased from 0 K, Einstein's value for the specific heat of a crystalline substance as a function of its temperature should approach the constant value at room temperature.

Einstein's predictions for the behavior of the specific heat of solids as a function of temperature were based on an application of the quantum hy-

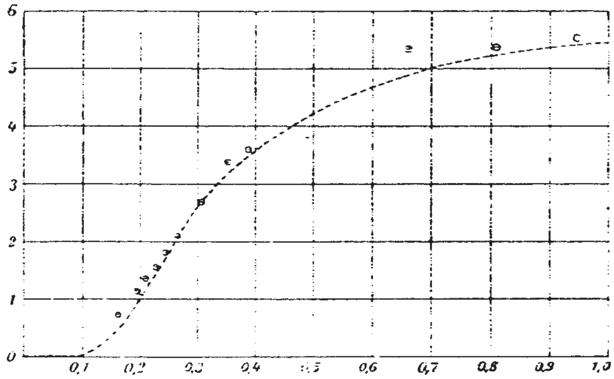


FIGURE 16.4 Einstein's graph of specific heat versus temperature from his 1907 paper applying the quantum hypothesis to Drude's classical oscillators.

pothesis to a model of a crystalline solid in which the lattice atoms behave as three-dimensional oscillators vibrating with quantized energies. The physicists Walther Nernst and Frederick A. Lindemann, working in a low-temperature laboratory in Berlin, astonished their colleagues with a complete confirmation of Einstein's prediction for the way the specific heat of a crystalline solid changed with temperature. It was the first confirmation of the quantum hypothesis outside of the field of radiation, and it clearly indicated that crystalline solids could be treated as an array of oscillating lattice atoms subject to requirements of quantum theory and, later, quantum mechanics. This model has served as the basic model of a crystalline solid ever since. As Nernst and Lindemann wrote, "That the observations in their totality provide a brilliant confirmation of the quantum theory of Planck and Einstein is obvious."

16.4 CONDUCTING ELECTRONS

You have seen in previous chapters how research on electricity and the behavior of electric currents enabled the enormous growth and spread of the electric age in which we live today. Electrical energy is now the dominant form of energy consumption in industrialized countries and is an essential element in the operation of many of the devices we use every day. For decades, one of the most poorly understood yet most practical properties of a piece of solid matter was its ability, or lack of ability, to conduct electricity. Why do some materials conduct electricity, while others act as insulators, and still others (such as semiconductors) act as one or the other depending upon the conditions? How can an understanding of such properties be put to even greater practical use?

You recall from Section 10.7 that the German physicist Georg Ohm discovered an empirical law, known as Ohm's law, relating the current in a material and the voltage applied to the material

$$V = IR.$$

Here V is the potential difference, I is the current, and R is the resistance of the material. The resistance is normally constant for each material, but there is a different resistance for different materials. As the resistance of a material increases, its conductivity goes down, so the amount of current that it conducts must also decrease, according to Ohm's law:

$$I = \frac{V}{R}.$$

Of course, if the resistance of a material is so large that it approaches infinitely, then V/R becomes zero, so $I = 0$. This means that no current flows through the material. Such a material is called an *insulator. An insulator allows virtually no current to get through*. On the other hand, if R ever happened to reach zero, the current would become infinite, or undefined, for even the tiniest voltage. This is what happens in a *superconductor*.

Even without a voltage source, a current in a loop made of a superconductor, once started, will continue in principle undisturbed virtually forever. The first superconductor was discovered in 1911 by Kammerlingh Onnes in Amsterdam while studying the conductivity of mercury at extremely low temperatures. During the late 1980s, one superconductor maintained a current without loss for nearly 3 years. A normal conductor would lose its current within a small fraction of a second without the help of an external voltage. Even with an applied external voltage, you know that a wire of normal resistance carrying a current begins to warm up, indicating that electrical energy is being converted into heat energy in the wire. Wires can become so hot that they glow and emit light; this is put to good use in incandescent light bulbs and in electrical heating elements. This behavior is similar to the conversion of kinetic energy into heat energy due to friction, as when a book slides across a tabletop and comes to rest. It seems that perhaps the current in a non-superconducting wire encounters a type of friction as it moves down the wire, which explains why the wire gets hot. This suggests that in an insulator there is so much internal friction that it does not allow any current to flow, while superconductors have no friction at all, allowing currents to flow indefinitely without any noticeable loss of energy.

Can we utilize the basic oscillator model of a solid to account for Ohm's law and the related electrical observations just described? Soon after Thomson's discovery of the electron, Drude in 1900 and H.A. Lorentz in 1909 developed the first electron model of conductivity in a conducting material. In their model, the solid material is assumed to consist of metal atoms arranged in a crystal structure, which is able to vibrate. However, electrons are the crucial element in understanding conductivity. Conducting metals such as copper and silver have one valence electron that is easily detached. So Drude and Lorentz assumed that these valence electrons provide the conduction current when a voltage is applied. Their model is called the *classical free-electron model*, because it involved only classical physics (no quantum theory) and because the electrons were considered to be little balls of charged matter that are completely free to wander about within the material. They were not subject to repulsive forces from other electrons or to attractions to the metal ions. Only a "boundary force" around the edge of the metal acted on the electrons, preventing them from escaping into space.

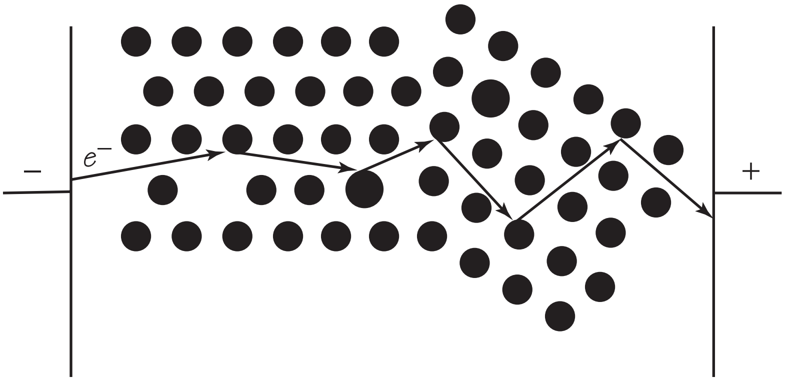


FIGURE 16.5 Schematic representation of an electron path through a conductor (containing vacancies, impurity atoms, and a grain boundary) under the influence of an electric field.

In many respects the electrons in the Drude–Lorentz free-electron model form a type of ideal gas, an "electron gas" inside the metal. Despite the presence of the lattice atoms, the electrons are assumed to move about randomly, like molecules in the kinetic theory of gases, until an electric field is applied. Since the electrons are charged, at the instant the external field is applied they start to move in the direction of the electric force on them, but as they move they do undergo inelastic collisions with the vibrations of the crystal lattice, as well as collisions with impurities in the material and with imperfections in the lattice. All of these work together to slow the electrons' advance and to reduce their kinetic energy. This slowing appears as a type of friction, which we surmised above, and the amount of "friction" is defined as the *resistance*. The lost kinetic energy appears as heat. The more inelastic collisions the electrons experience as they traverse the conductor, the more the resistance they encounter; on the macroscopic scale this is expressed as Ohm's law.

Of course, if the material is made up of atoms that have no valence electrons (which can be shared), such as glass, wood, or plastic, there can be no conduction of electrons, and thus no current when a voltage is applied. The material is an *insulator*.

The Drude–Lorentz free-electron model accounted quite well for Ohm's law, but it could not account for some of the thermal properties of the metals, besides the specific heat, such as the dependence of the resistance upon the temperature. Moreover, with the advent of quantum mechanics it was obvious that a quantum mechanical understanding of conductivity was required, since atoms and electrons obey the laws of quantum mechanics.

16.5 BANDING TOGETHER*

In 1928, just two years after the formulation of quantum mechanics, the German physicist Arnold Sommerfeld modified the classical free-electron model by treating the electrons according to quantum mechanics. But the new theory still contained the unrealistic assumption that the electrons do not interact with the charged lattice icons except to collide with them. As before, Sommerfeld also considered the electrons to be little charged particles of matter.

Beginning in the same year, Felix Bloch, an assistant to Werner Heisenberg in Leipzig, began to make more realistic assumptions in an attempt to formulate a more complete quantum mechanics of electrical conductivity. First, because he wanted to assign a definite momentum and energy to each of the electrons, but not a definite position or a time interval, he chose the wave side of the wave-particle duality. He assumed that the electrons behave, not like particles, but like infinitely extended de Broglie waves. As a result, Bloch did not treat electrons inside metals as a "gas" of particles, but rather as periodic waves extending throughout the periodic crystal lattice. This, it later turned out, helped to explain how electricity can begin to flow in a wire the instant a wire is plugged into a wall socket. If the electrons are viewed as balls of matter, it would take a small amount of time for the current to begin flowing at the rate specified by Ohm's law.

Bloch made a second assumption. He assumed that the positive metal ions, which are arranged in an infinite, periodic array (that is, in a perfect crystal), each exerts an attractive electric force on the negative electrons. This attractive force formed in visual terms a potential energy that looked like a type of "potential well." The wells of neighboring ions then overlapped so that together they formed a periodic arrangement that gave the electron waves a very bumpy ride down the wire.

Bloch then solved the Schrödinger equation for the energies that these types of de Broglie waves (wave functions) could possess while moving in this type of periodic potential. He discovered that the allowed energies of the electrons in the material are joined together into *bands* of quantum states, just as there are certain quantum stationary states within each atom in which the electrons are allowed to exist. Between the bands, as between the quantum states, there is a range of energies in which electrons are forbidden to exist. The bands in the material are actually created by the join-

---

* The remainder of this chapter contains material that is somewhat more advanced than that in the previous sections. Your instructor will inform you how much, if any, of this material is included in the course work or in individual assignments. You may find some of it of interest for independent study.

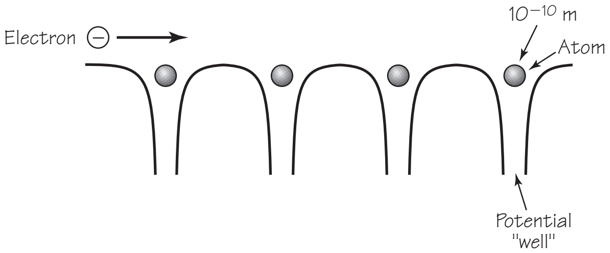


FIGURE 16.6 An electron moving through a region of potential wells.

ing together of the quantum states of the individual atoms. In fact, if there are a total of N identical atoms in the material, then there are N quantum states within each band. According to a rule in quantum mechanics (the Pauli exclusion principle), only two electrons are allowed to occupy any one quantum energy state of a single atom, and this is allowed only because the two electrons spin on their axes in opposite directions.

As extended and refined by Bloch and others during the 1930s, Bloch's theory, known as the *band theory* of metals, accounts very well for the conducting behavior of materials. When atoms are joined together into a crystal, each of the individual quantum states of the atoms joins with the corresponding states in other (identical) atoms in the crystal to form the various energy bands within the material. The electrons in the atoms then fill up the available states within each band. For instance, if there are N atoms in a material, there are N states in a single band and up to $2N$ electrons in each band (since two electrons can occupy one energy state if they spin in opposite directions).

The interesting physics occurs at the top of the filled bands. When an external electric field (voltage) is applied to the material, an electron can respond to the field only if it can move up to a slightly higher quantum state, since it would have slightly more energy in responding to the field.

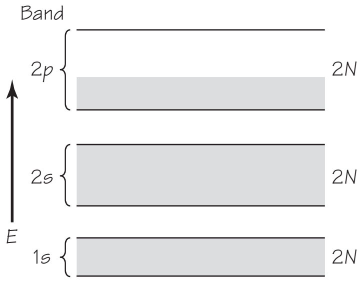


FIGURE 16.7 Expected occupation of the energy bands in a crystal of N atoms with $6N$ electrons. $2N$ electrons go into each band, filling the first two but leaving the third band partly empty.

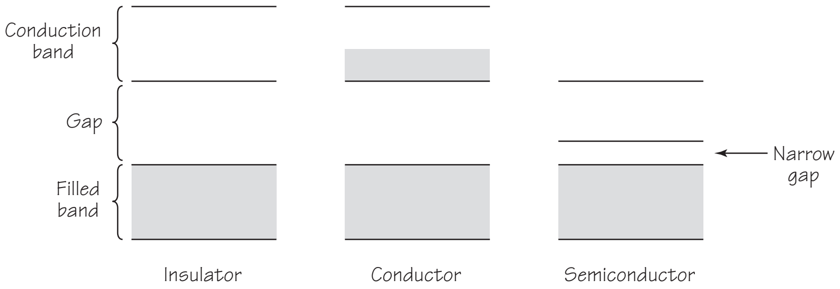


FIGURE 16.8 Bands for insulator, conductor, and semiconductor.

This is possible only if there are some nearby free states to which the electron can jump. This is the case for conducting metals, such as copper and silver, since as their electrons fill in the available states, the highest band is only partially filled. The electrons in this band, which is called the *conduction band*, are then free to be conducted (actually, propagated as waves) down the wire, since there are empty quantum states nearby in their energy band. But the electrons in the filled lower bands are not able to move, since there are no free states nearby.

On the other hand, if the filling of the states within the bands stops exactly at the top of a band, then the electrons cannot respond normally to an electric field by moving down the wire, since there is an energy gap preventing them from reaching a quantum state in which they can move freely. Such a material would then be an insulator. An example is diamond, in which the highest filled quantum state also fills out an energy band. This band is called the *valence band*, since it is occupied by the outer valence electrons of the atoms. The energy gap to the next band, which is empty,

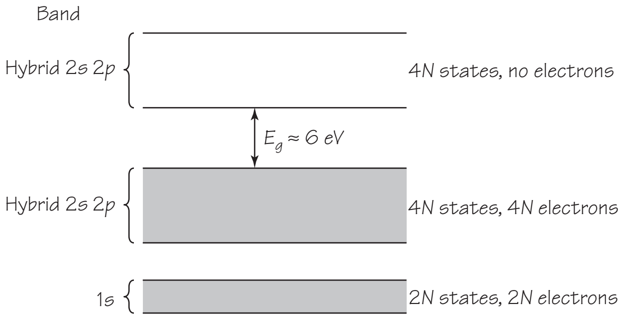


FIGURE 16.9 Actual occupation of the energy bands in a diamond crystal with N atoms ($6N$ electrons). The two lower bands are full and separated by an energy gap Eg from the higher hybrid *2s2p* band, which has no electrons.

is about 6 eV, so the electrons in diamond cannot normally reach the conduction band and diamond is therefore a good insulator.

16.6 SUPERCONDUCTORS

Good conductors usually cannot become superconductors.

The band theory of metals has been subjected to experimental tests many times and is now the accepted theory of the behavior of conductors and insulators. But it has to be modified when it comes to superconductivity. We noted earlier that electrical resistance is due to collisions of the electrons (whether treated as particles or waves) with impurities, imperfections, and especially the lattice vibrations of the metal crystal. The lattice vibrations of the solid will decrease as the temperature falls, because the entropy, which represents disorder, also decreases. Therefore the resistance should also decrease. This is what is observed, and it is well explained by quantum band theory. But at very low temperatures something strange happens: the resistance of some materials suddenly drops to zero at a certain temperature. As noted before, these materials become *superconductors*, meaning they can conduct currents without resistance or loss of electric energy. The temperature at which superconductivity occurs is called the *critical temperature*. It ranges from 0.015 K for tungsten to as high as 160 K for certain ceramics.

You can imagine some of the possible advantages of electrical wires having zero resistance. It is like motion without friction. Among the many actual and potential applications are electric power transmission lines without power loss, faster computers, and super-strong electromagnets. The latter are used today in magnetic resonance imaging (MRI) devices, high-energy particle accelerators, and high-speed levitation trains (often called "mag-lev" for magnetic levitation). Since the boiling temperature of liquid nitrogen, which is plentiful and relatively easy to produce, is 77 K, superconductors with transition temperatures higher than 77 K can be kept suf-

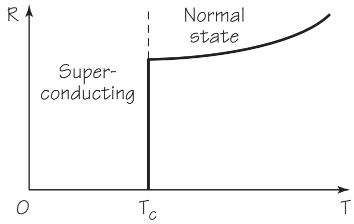


FIGURE 16.10 Transition of resistance (R) to superconductivity at the critical temperature T_c.

FIGURE 16.11 Japanese Mag-Lev train driven by superconducting electromagnets.

ficiently cold by keeping them immersed in liquid nitrogen. However, such materials are brittle, and the superconductivity can be easily destroyed by stray magnetic fields. Researchers today are working to overcome these problems and to create superconducting materials with transition temperatures even closer to room temperature (20°C, or 293 K).

The development of superconducting devices was greatly stimulated after the acceptance of the basic theory of superconductivity proposed in 1957 by John Bardeen, Leon Cooper, and Robert Schrieffer. The authors of the BCS theory, as it is known, received the Nobel Prize for their work in 1972. Their theory is highly technical but the basic idea is that the electron waves in the superconducting state no longer act independently, as in Bloch's model. Instead, they are paired together at the critical temperature so that their wave functions act as one unit as they interact with the crystal lattice. Moreover, all of the electron pairs move together in one collective motion, so that if any single electron is scattered by the lattice it is pulled back into the flow by its partner, and if any pair of electrons is somehow scattered off track, it is pulled back into the collective flow by all of the other pairs. Since there is no scattering or inelastic collisions, there is no resistance, and the material becomes a superconductor.

Although the BCS theory accounted well in general for certain crystalline solids, further refinements are required today for other materials, such as ceramics, and for the more detailed behavior of superconductors.

16.7 SEMICONDUCTORS

The most prevalent use today of semiconductors is forming them into transistors, the basic electronic building blocks of all “solid-state electronics” and computer microchips. Semiconductors are the most important materials in the burgeoning revolution in computers and other electronic devices today. Such applications also arose, beginning in the 1930s, from the quantum mechanics of the band structure of solids.

The most common semiconductors are made of silicon or germanium, elements 14 and 32 on the periodic table. It was known that they form very stable crystal structures that should be insulators but are in fact weak conductors of electricity. For both of these elements, it was found that the numbers of electrons are just enough to fill up to the top of an energy band in each case. This is why they should be insulators, and, in fact, at near absolute zero, 0 K, they are insulators (not superconductors).

At very low temperature, the lattice vibrations in silicon and germanium are minimal, and the electrons at the top of the valence band are not able to obtain enough energy from the lattice vibrations to enable them to jump the energy gap to the next band and become conducting. However, the gap to the next band is very small, only 0.7 eV for germanium and only 1.1 eV for silicon. Because these gaps are so small, at temperatures somewhat above absolute zero the electrons can pick up enough energy from the vibrations of the crystal lattice to jump the gap and land in the empty conduction band. So at room temperature these metals, which by their structure should be insulators, are actually weak conductors.

As failed insulators and poor conductors, silicon and germanium were not much use in electronics until the 1950s when advances were made in

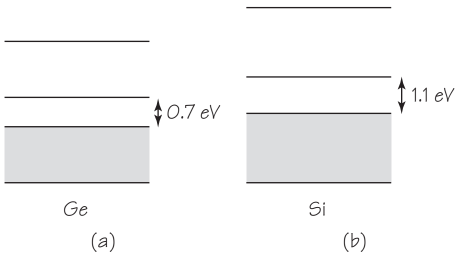
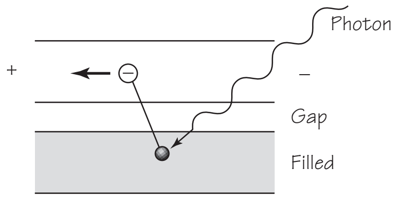


FIGURE 16.12 Energy bands of germanium (Ge) and silicon (Si).

FIGURE 16.13 The photoelectric effect of an electron in a semiconductor jumping into the conduction band with a potential difference across the material.



the controlled introduction of certain impurities into the lattice structure (see Section 16.8 below). But the real use of these two plentiful metals took off during the 1980s with the introduction of mass-production methods for super-thin, microscopically structured layers of silicon and (to a lesser extent) germanium crystals, which, when properly arranged (as also discussed below) can act as transistors. Today, wafer-thin layers of silicon, when made into micro-transistors through the introduction of impurities and broken up into "chips," are the basis of the trillion-dollar-per-year electronics and computer industries, and there is as yet no end in sight to the revolution they have unleashed. And it all depends on the narrow energy gaps in these crystals.

Since germanium and silicon are so sensitive to impurities, their large-scale use as semiconductors did not occur until methods were developed for producing ultra-pure graphite for nuclear reactors and ultra-pure germanium for electronic circuits during World War II. Pure germanium was also used at first in photoelectric cells. A photon from the outside can strike an electron in the valence band of germanium (and later silicon), providing the electron with enough energy to reach the conduction band in a type of internal photoelectric effect. For this to occur, the energy of the photon must be at least 0.7 eV for germanium and 1.1 eV for silicon. From the relationship for the energy of a photon, $E = hf$, these energies correspond to photons possessing frequencies in the infrared range of electromagnetic waves. Any waves (or photons) with frequencies in the infrared or higher range, which includes visible light, will cause electrons to jump into the conduction band and form a current. This is one reason these crystals are known as *semiconductors*, since they are good conductors only when the band gap can be overcome.

This type of light-induced conductivity can be used in a *photoelectric cell*, or *photocell* for short—a cell that produces electricity when light shines on it. You can imagine some of the many possible applications of a photocell. It can be used, for instance, for motion detectors. A light beam shining on a photocell in a circuit will generate a steady current. If somebody walks through the beam, this will interrupt the current, which could either set off an alarm or open an automatic door for the person to exit or enter. Automatic controls on night lights use the same principle. Since photocells are sensitive even to infrared rays, as long as there is sufficient daylight the cell will produce a current. When the sun goes down, the current stops, which signals a circuit to switch on the lights.

16.8 INTRODUCING IMPURITIES

The most extensive use of semiconductors, such as silicon or germanium, including their use as transistors, arises from their behavior when, after being sufficiently purified of atoms other than the basic element (e.g., silicon or germanium), very small amounts of special impurities are carefully introduced. While the methods for first purifying then adding small amounts of impurities to germanium became available after World War II, research on the purification and controlled introduction of impurities for silicon did not succeed until the late 1950s. Since silicon is more abundant than germanium and other semiconductors, it soon replaced germanium as the preferred semiconductor.

Germanium (element 32) and silicon (element 14) each have four valence electrons, which fill up the valence band when these metals form a crystal. Element 33, arsenic, has five valence electrons, as does element 15, phosphorus. If a very small amount of arsenic is added to germanium as the germanium crystal is formed, the arsenic atoms will substitute for a germanium atom in the lattice. The same happens when a small amount of phosphorus is added to silicon. The small amounts are carefully controlled in the production process and are only about 0.0001% of the total atoms. This process is calling *doping*, and the impurities are called *dopants*.

When the corresponding dopant is introduced into the germanium or silicon lattice, four of its valence electrons are bonded to the other atoms in the lattice, leaving the fifth valence electron without a bond. These extra valence electrons from the dopants are so lightly attached to their atoms that they easily absorb vibration energy and jump into the conduction band

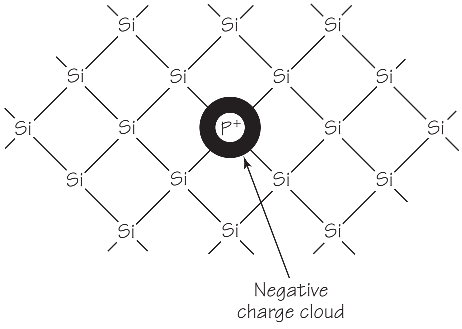


FIGURE 16.14 Two-dimensional representation of a silicon lattice in which a phosphorus atom substitutes a regular lattice atom, and thus introduces a negative charge cloud about the phosphorus atom. Each electron pair between two silicon atoms constitutes a covalent bond.

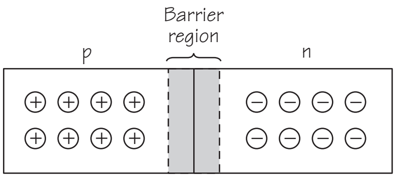
of the semiconductor. At room temperature they are all in the conduction band, which means that the doped semiconductor now acts like a full conductor. Since the negative electrons donated by the impurities enable the material to conduct, these types of semiconductors are called *n-type semiconductors*.

A similar situation occurs when the semiconductors are doped with impurities one space lower on the periodic table, that is when silicon (14) is doped with element 13, aluminum, and germanium (32) is doped with element 31, gallium. The dopants have only three valence electrons. So, when they are bound into the silicon or germanium lattice, they leave a space in the valence band indicating that they are able to accept an electron from a neighboring silicon or germanium atom. This space is called a *hole*. The hole behaves similarly to an air bubble in a glass of soft drink or beer. When an external electric field is turned on, electrons from the neighboring atoms, which normally cannot move anywhere, can now move over one atom in filling up the hole. But then this leaves a hole in the neighboring atom, which can be filled up by its neighbor, and so on. Similarly, the upward motion of the air bubble in a glass of beer actually involves the downward flow of beer into the space occupied by the bubble, which appears to the eye to be moving upward. In the semiconductor, if the field moves the electrons to the left in the diagram, the hole appears to migrate to the right. By moving to the right, it is behaving like a positive charge, even though it is just a space. Nevertheless, because of the motion of negative charge to the left, the physical effect is the same as if positive charges were flowing to the right in the valence band, below the conduction band, where previously no charges could flow at all. Once again, the very careful introduction of small amounts of impurities has turned the semiconductor into a full conductor. Since the conducting charges appear to be positive, this is known as a *p-type semiconductor*.

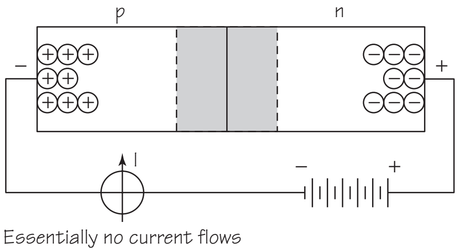
These doped semiconductors could be used like any other conductor, but that is hardly worthwhile, since we can produce full conductors much more easily and cheaply, for example, copper wires. Instead, it was recognized during and after WWII that the really useful applications of these types of semiconductors is when they are placed physically next to each other inside electronic devices.

16.9 SEMICONDUCTOR DEVICES

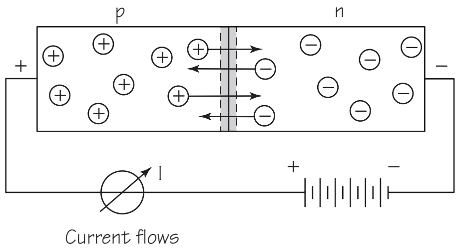
Let's take a p-type semiconductor with a very clean surface and place it right next to the clean surface of a n-type semiconductor. This type of device is called an *n-p junction diode*, or simply a *diode*. What happens? Scien-



(a) No bias



(b) Reverse bias



(c) Forward bias

FIGURE 16.15 An *n-p* diode with (a) no bias; (b) reverse bias; and (c) forward bias.

tists realized that the electrons in the conduction band of the n-type semiconductor are able to move across the boundary between the two semiconductors, and fall into the holes in the lower energy valence band of the p-type semiconductor. The electrons and holes begin to disappear. But as the electrons disappear in the n-type, the positive charges of the impurities are no longer balanced by the negative conducting electrons, so the n-type becomes positively charged. The opposite happens in the p-type semiconductor. As the holes disappear, the impurities are left holding an extra electron, which causes the p-side to become negatively charged. The result of all this is that after a very short time a net electric field is set up between the n-type and p-type semiconductors which stops the process of mutual destruction by keeping the electrons on the n-type side separated from the holes on the p-type side, with a "depletion layer" in the middle.

Now, if a potential difference, such as one caused by a small battery, is placed across the device from the p-type side to the n-type side, the equilibrium may or may not be disturbed, depending upon how the potential is placed. If the positive wire is placed on the n-type material and the negative wire on the p-type material, as in Figure 16.15(b) the separation of the electrons and holes is reinforced. So no current will flow. But, if the wires are reversed, so that the positive wire is placed on the p-type material and the negative wire on the n-type side, the negative electrons on the n-type side will be pushed toward the boundary, and the same for the positive holes on the p-type side. If the external potential is greater than the internal potential of these two, a current will flow in the device.

The current will be enhanced if now a photon arrives from the outside and pushes an electron from the n-type material into the conduction band. The n-p diode can then be used as an even more powerful photocell than a simple semiconductor alone. Perhaps the most common use of this type of photocell is in the generation of electricity from solar energy. This can be done for small-scale devices, such as pocket-calculators, or for larger scale energy needs, such as the energy for an entire building. Such photocells are often known as *photovoltaic cells*. The International Space Station and other satellites near the earth's orbit depend for their energy needs upon huge "wings" of photovoltaic cells, unfurled to catch the radiation from the sun. Research continues on how to make even greater use of photoconductivity from the sun's rays to produce electrical energy for our needs on a large scale.

Deep-space satellites use energy from isotope decay, since they are too far from the sun for solar panels to be of much use.

FIGURE 16.16 Artist's conception of International Space Station with its "wings" of photovoltaic cells.

A photocell working in reverse is also useful. If a conduction electron in an n-p diode happens to fall into a hole, it will emit the excess energy as a photon, much as an electron would do in an atom as it jumps to a lower quantum state. If the diode is designed to emit light in the visible range for display purposes, it is called a *light-emitting diode*, or *LED* for short. These LED lights, which are usually red, green, or orange, can be found on most electronic devices, usually to indicate that the device is on or that the circuit is functioning. They can be also used in large-scale devices, such as computer monitors, where they consume less energy and exhibit less distortion than do traditional CRT monitors.

One of the biggest uses of n-p diodes is in circuits where we want the current to flow in one direction only, but not in the opposite direction. This is used, for instance, in logic circuits for computers, in which an answer of “true” or “false,” or “yes” or “no,” or 0 or 1 can be decided by whether or not the current is allowed to flow through the device. A device that allows a current to flow in one direction only can also be used to convert an alternating current (AC), for instance the current from a wall socket, into direct current (DC) for use in small electronics devices. Commercial electric current in the United States alternates in a wave-like pattern at a rate or frequency of 60 cycles per second (60 Hz). This is shown in Figure 16.17a. If we hook up a closed circuit that includes the AC wall socket, a device for measuring the current (an ammeter), and an n-p semiconductor, the current in the wire will go in only one direction, which occurs only when the voltage is positive on the p-type side and negative on the n-type side. This conversion of AC to DC current is called *rectification*, and devices that rectify are naturally called *rectifiers*. These are useful for devices that accept only DC current.

You can see from Figure 16.17b that the voltage produced by the n-p junction diode is positive only, but it is constantly changing over each hump

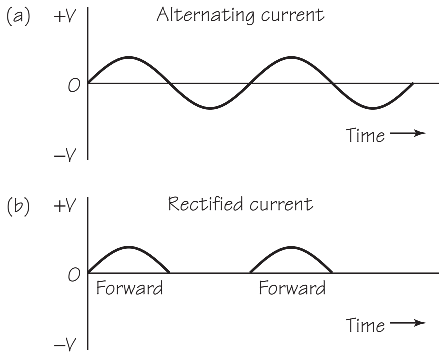


FIGURE 16.17 Graphs of AC and rectified voltages resulting in AC and DC currents.

and hits zero for a long interval. These effects can be reduced by sending the current through additional electronic devices, known as *filters*, that help to smooth out the voltage to a constant value, producing a steady direct current that can then be fed into a DC electronic device—an electronic game, a calculator, a laptop computer, etc.

These properties of semiconductors become even more interesting when we add a third semiconductor to our junction diode. This forms a *transistor*.

16.10 TRANSISTORS

The properties of n-type and p-type semiconductors described above were well known by the end of World War II, during which considerable research was devoted to electronics and the invention of radar. In 1947 three researchers at Bell Laboratories in New Jersey took the idea of the n-p junction diode a step further. William Shockley, Walter Brattain, and John Bardeen (who, alone of the three, later went on to work on the theory of superconductivity), placed two germanium diodes back to back—an n-p diode next to a p-n diode, each sharing the p-type semiconductor. They found that the device could be made to contact electricity or to resist, that is, to block, the flow of electricity, depending upon the charge applied to the middle layer of p-type semiconductor. Such a device is therefore called a *transistor*, since it transmits or resists current depending upon the charge provided to the middle layer. The type of transistor just described is called a *n-p-n bipolar transistor*. A *p-n-p bipolar transistor* is also viable. (There is also a different type of transistor construction, known as the "field effect transistor," which works on the same general principle.)

By the early 1960s, as silicon-based transistors had become smaller, more reliable, and cheaper to make, they began replacing the bulkier and more expensive vacuum tubes previously used for such purposes as amplifying voltages, for instance in stereo sound systems, TVs, and radios, and as switching and logic devices in computers. Shockley, Bardeen, and Brattain received the Nobel Prize in 1956 for an invention that would soon launch the revolution in computer electronics—the transistor. Shockley, based at Stanford University, went on to lay the foundations for the development of Silicon Valley industries in the vicinity of the university. Many of the companies there, including Intel, were founded by people who had originally worked with Shockley.

The revolution in computer electronics was hastened when the need for miniaturized electronic components for space-age missiles helped lead to

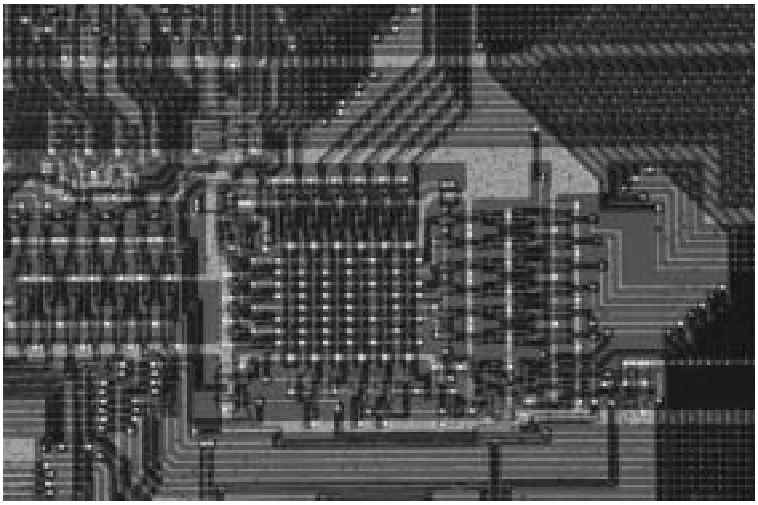


FIGURE 16.18 Close-up of a portion of a semiconductor chip.

the development of the *integrated circuit*. In 1958 Jack Kilby of Texas Instruments developed the idea of integrating transistors and related circuitry on a single silicon chip. A year later, Robert Noyce and Jean Noerni of Fairchild Semiconductor Corporation put the idea into practice by developing the method for integrating the separate elements onto the chip. In 1970 reduction in the size of integrated circuits was introduced with the first microprocessor, and in 1971 the newly formed Intel Corporation introduced the first integrated microprocessor circuit. Continued advances in microprocessor design and manufacturing techniques enabled the mass production of microprocessors for personal computers, cell phones, automobiles, industrial robots, and even toasters and “talking” dolls. By the early 2000s Intel and its competitors could place as many as 42 million microscopic transistors on a single silicon chip, two centimeters square.

Figure 16.19(a) shows a schematic diagram of an n-p-n bipolar transistor with a voltage placed across it. Figure 16.19(b) shows how such a device is actually constructed by “chip makers” such as Intel and Advanced Micro Devices. A thin layer of n-type silicon is placed over a thin “chip” (crystal) of very lightly doped p-type silicon. The p-type silicon, which is sandwiched between the n-type pieces, is called the “base.” Being very thin and only lightly doped, most of its electrons, which will move through it from left to right, will not be captured by the holes, yet they will still provide an internal field to prevent any flow when there is no external field. The entire chip must also be very thin in order to allow the electronic cir-

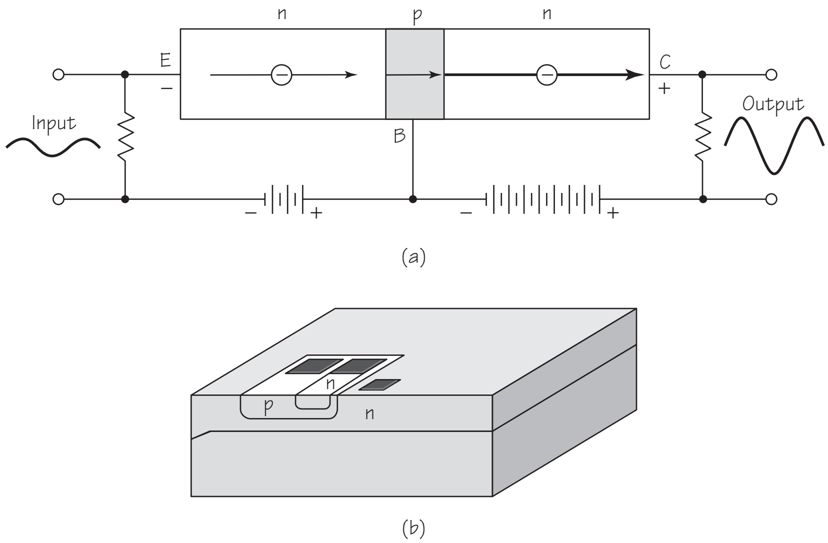


FIGURE 16.19 Schematic representation of an *n-p-n* bipolar transistor (see text). The dark areas in (b) are the contact pads.

cuitry to function as rapidly as possible. Using a photographic emulsion, a microscopic circuit pattern is placed on the surface layer of the n-type silicon such that the n-type surface layer is stripped away at certain places when an acid solution is applied to the surface. This exposes the p-type silicon underneath, so that the electrical connections can be made to form the n-p-n transistor.

What does the transistor actually do, and how does it do it? Think of Figure 16.19(a) as representing two n-p diodes back to back. Remember that if a negative or zero charged wire is placed on the p-type silicon and a positive charged wire is placed on the n-type side, no charge will flow. However, if a positive wire is placed on the base (p-type) and a negative wire on the left n-type silicon, charges will flow from the n-type silicon into the base. If the base is very thin and only lightly doped, the charges will make it through the base into the n-type silicon on the right. Here the situation is reversed. The p-type stays positive, but an even higher voltage positive wire is attached to the end of the right-hand n-type silicon. The purpose of this is to keep the electrons moving from left to right. In addition to the conduction electrons that are already in the right-hand n-type silicon, new electrons are arriving from the n-type and p-type silicon on the left. The amount of flow can be controlled by the positive voltage at the base. The net effect is to amplify (increase) the current from left to right.

16.11 SOME APPLICATIONS OF TRANSISTORS

Amplifying Voltage

If we place a large resistor in a circuit containing a transistor, as in Figure 16.19(a), we can amplify the incoming voltage, since according to Ohm's law voltage is equal to current times resistance. So a current crossing a small resistance corresponds to a small voltage. But if the same current is made to cross a much larger resistance, it will yield a much larger voltage, that is, the small input voltage will be amplified to yield a large output voltage.

For example, the amplified voltage can correspond to a portion of a digital sound wave, in which the analog wave has been approximated by a series of single voltages. This small portion of the wave (one bit) can be represented by a tiny voltage received at the transistor base from a photocell that is stimulated by a laser beam that picked up the signal from a music compact disc (CD). The amplified voltage from the transistor can then be sent to a speaker set, which turns the signal into a vibration that is emitted by the speakers as a pressure wave that your ears detect and your brain interprets as—let's say—the beautiful strains of Beethoven's *Symphonie Pastorale!*

DRAM

As a logic device, a transistor can act like a diode in allowing current to pass or in preventing current flow, depending upon the voltage at the base. By combining p-n-p and n-p-n transistors together, along with diodes, various logic systems and "gates" can be created. One of the most useful

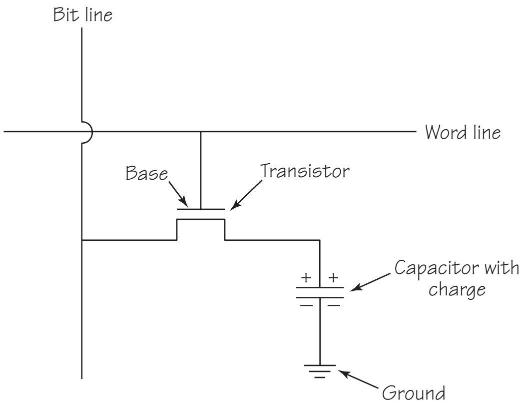


FIGURE 16.20 Schematic representation of a DRAM cell.

applications of this is in the creation of "dynamic random access memory" (DRAM) in computers, devices that store data that can be accessed at will in random order. As is well known, information is stored digitally in computers, that is, it is stored in the form of strings of binary numbers made up of just 1s and 0s. Each binary number (whether it is 1 or 0) is called a "bit." (Eight bits make a "byte.") For instance, the binary number 10101010, which has eight bits (making one byte), is equivalent in the decimal system to the number 170. If you happen to type the number 170 onto a spread sheet and save it, the computer will first convert the decimal number 170 to the binary number 10101010. It will then assign each of the eight bits to a single cell within dynamic random access memory. Each bit will correspond to the charge on a charge device known as a capacitor, which, in principle, is just two plates held next to each other. If the bit is 1, the capacitor will receive a small charge. If the number is 0, it receives no charge. One of these cells is shown in Figure 16.20 with the number 1 on it (plus charge on the capacitor).

Now you ask the computer to retrieve the number 170 from memory. Here's what happens. Your request is sent to the "word line" as a positive voltage, which is then applied to the base of the transistor in each circuit on that line. Since the base is now positive, the charge on the capacitor can flow through the transistor and up through the "bit line" to the processor which detects the current and identifies the number in that cell as a 1. Of course, if there is no charge stored in the capacitor to begin with, there will be no current on the bit line, and that bit will be identified as 0. Figure 16.21 shows what the cells for all eight bits would look like for the decimal number 170. This diagram represents only one byte of memory. Many personal computers today now come with hundreds of "megabytes" (MB), or even gigabytes of memory, miniaturized on microprocessor memory modules. For example, 128 million bytes of memory, or $8 \times 128 \times 10^6$ bits, are 1024×10^6, or about *one billion cells* like Figure 16.21! It is aston-

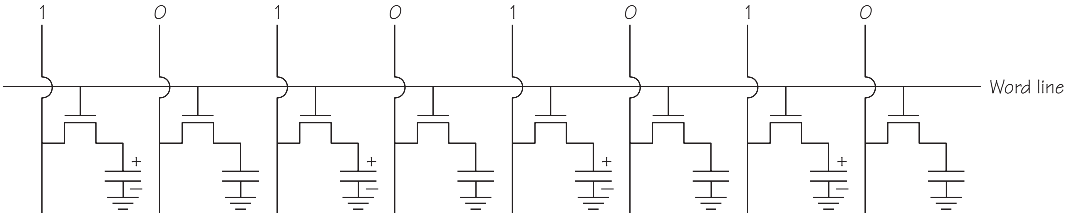


FIGURE 16.21 One byte of DRAM showing the decimal number 170 as the binary number 10101010. The 1 is indicated by the charge on the capacitor. 0 is indicated by no charge on the capacitor.

ishing how far this technology has come in only a few decades since the invention of the transistor in 1947. Where will we be in another 50, or even another 10 years?

SOME NEW IDEAS AND CONCEPTS

amorphous solids
band theory
bit
byte
classical free-electron model
composite solids
conduction band
critical temperature
crystalline solids
diode
dopant
doping
hole
insulator
latent heat of fusion
LED
n-type semiconductor
p-type semiconductor
phase
photocell
photovoltaic cell
quasi-crystalline solids
rectifier
semiconductor
superconductor
transistor
valence band

FURTHER READING

I. Amato, *Stuff: the Materials the World is Made Of* (New York: Avon, 1997).
P. Ball, *Made to Measure: New Materials for the 21st Century* (Princeton, NJ: Princeton University Press, 1997).
M. Campbell-Kelly and W. Aspray, *Computer: A History of the Information Machine*. Sloan Technology Series (New York: Basic Books, 1997).
G. Holton, H. Chang, and E. Jurkowitz, How a Scientific Discovery Is Made: A Case History [High-Temperature Superconductivity], *American Scientist*, **84** (1996), 364–375.
R.E. Hummel, *Understanding Materials Science: History, Properties, Applications* (New York: Springer-Verlag, 1998).
M. Riordan and L. Hoddeson, *Crystal Fire: The Birth of the Information Age* (New York: Norton, 1997).

STUDY GUIDE QUESTIONS

16.1 The Success of Quantum Mechanics

1. Describe some of the successes of quantum mechanics.

16.2 Forming a Solid

1. What happens to the temperature of a material at a phase transition, and why?
2. A container of cold water is placed in a refrigerator. The refrigerator is constantly running but the reading on a thermometer in the water remains unchanged. What do you conclude is happening to the water?
3. What exactly happens when a liquid freezes into a solid?
4. Why would the lowest entropy state for a solid involve the largest number of possible bonds?
5. What is the most prominent characteristic of a crystal structure?
6. Describe the other three main types of solid bonding.

16.3 Quantum Solids

1. What did Einstein first show about the charges in solids?
2. Describe Drude's oscillator model of solids.
3. In what ways did Einstein's model go beyond Drude's?
4. What did Einstein predict in 1907 from his model?
5. Was his prediction confirmed, and why was it so significant?

16.4 Conducting Electrons

1. Using Ohm's law, describe how insulators and conductors differ from each other.
2. Describe the classical free-electron model of a solid.
3. How did the classical model account for conductivity and Ohm's law?
4. Evaluate the classical model as a theory.

16.5 Banding Together

1. How did Sommerfeld modify the classical free-electron model of solids?
2. What assumptions did Bloch make in constructing a quantum model of metals?
3. What did Bloch discover when he solved Schrödinger's equation?
4. Describe the band theory of metals. How are the bands formed?
5. What is the difference between an insulator and an conductor, according to the band theory?

16.6 Superconductors

1. What is superconductivity, and where does it occur?
2. Scientists are now trying to obtain superconducting materials as close as possible to room temperature. What would be some of the advantages to this?
3. Briefly, what is the BCS theory of superconductivity?

16.7 Semiconductors

1. What is a semiconductor?
2. According to the band theory, why are semiconductors "failed insulators and poor conductors"?
3. How might semiconductors be used as photodetectors?

16.8 Introducing Impurities

1. How is an n-type semiconductor formed?
2. What is it that conducts in an n-type semiconductor?
3. How is a p-type semiconductor formed?
4. What is a “hole,” and why is considered to be positive if it is really just an empty orbit on an ion?

16.9 Semiconductor Devices

1. What is an n-p junction diode?
2. How does an n-p junction diode reach equilibrium?
3. How does a photovoltaic cell work?
4. How does a LED light work?
5. How might a diode be used as a rectifier?

16.10 Transistors

1. How is a transistor made from semiconductors?
2. What is an integrated circuit?
3. How are millions of microscopic transistors placed on a single silicon chip?
4. What does a transistor actually do and how does it do it?

16.11 Some Applications of Transistors

1. How might a transistor be used as an amplifier?
2. How might a transistor be used to store information in a computer?

DISCOVERY QUESTIONS

1. How do the free-electron theory and the band theory each account for the existence of insulators?
2. Heat is added to a melting ice cube until the cube has completely melted. The heat that goes into the cube is called the “latent heat of fusion,” the word *latent* meaning “hidden.” Where is the heat hidden?
3. Why is it that the temperature of a melting ice cube and its melted water stay at 0°C until all of the ice has melted?
4. Why can’t some materials remain in the liquid state all the way down to absolute zero?
5. Why is it that some elements, such as mercury, are liquid at room temperature?
6. Whenever you plug an appliance into a wall socket the current flows instantly. Should it not take some time for the electrons to begin to flow down the wire?
7. Why did Bloch decide to view the electrons as waves rather than as particles?
8. According to the band theory, how do an insulator, a conductor, and a semiconductor differ from each other?
9. What are some of the applications of semiconductors?

10. Why do impurities have to be added to semiconductors to make them into conductors?

Quantitative

1. Translate the binary number 100 into a decimal number.
2. Translate the decimal number 100 into the binary system.
3. What would a memory circuit look like that is storing the decimal number 100 in binary form?