CHAPTER 15

Atomic Physics Pays Off: Solar Cells, Transistors, and the Silicon Age

Solid-state devices, of which the transistor is the most important, have had an enormous impact on our civilization. Their development was made possible by our knowledge of atomic structure and of the motion of electrons in solid materials. The understanding of the process of electric conduction is one of the triumphs of twentieth-century physics. It depends crucially on the quantum mechanics first developed by Schrödinger and Heisenberg in 1925. The invention and proliferation of the various types of transistors are unthinkable without this theoretical development. But it could also not have happened without the experimental advances in materials science that allowed the production of crystals of silicon and other elements with extreme purity or with impurities that are precisely controlled in quantity and spatial distribution. As so often in history, theory and experiment complemented each other as they progressed hand in hand.

15.1 The real solid: metals, insulators, and semiconductors

Electrons in a metal: not so free

The understanding of electric conduction and of metallic behavior became possible only after the development of quantum mechanics. A part of that triumph with the most profound influence was the understanding of why some materials are metals and some are not, and the control over conduction properties that then became possible. The term *semiconductor* implies a conductivity between that of a metal and that of an insulator, but that doesn't begin to deal with the essential differences between these three classes of materials. To get further we have to go beyond the free-electron model of metals.

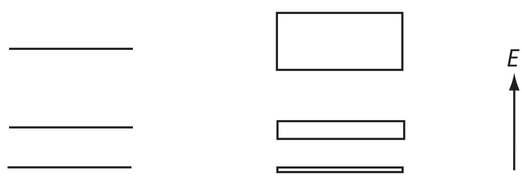
Of course the electrons in a metal are not free. First of all they are held to the metal in accord with Coulomb's law so as to preserve the equality of positive and negative charges. You might think that they have to twist and turn past the atoms, which are now ions, and that this is the greatest impediment to their motion. This is not the case. In fact, a rigid symmetrical arrangement of ions, a *perfect lattice,* does not impede the electron motion at all, a fact that can be understood and makes sense only as a consequence of the wave nature of the electrons.

The reason is the same as that which allows light to go through a window. The question needs to be turned around: not "why does the light wave go through?" but "what prevents it from going through?" If it does not go through, if it is *absorbed*, there must be some mechanism by which it loses energy. The energy of the photons, E_{ph}, must be just equal to an energy that can be accepted by the material. It must be equal to the energy difference, ΔE, between an energy level that is "occupied" and a higher one that is not. The atom, or the whole solid, can then go from the lower level to the higher one. The photon disappears. It is absorbed. If there are no energy levels with the required difference in energy, the photons will pass through without being absorbed, and the material will be *transparent* to the light. Glass, for example, has no energy levels that allow the absorption of visible light, but it absorbs light in the ultraviolet part of the spectrum. The story is the same whether we talk about light going through a window or electrons going through a metal.

The figure shows two levels separated by the energy difference ΔE. If the lower level is occupied, the photon with just the right energy, $E_{\text{ph}} = \Delta E$, will be absorbed. But the system with these energy levels is transparent to other photons.

Bands and gaps

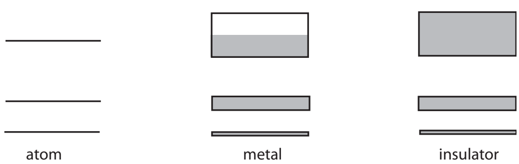
How is it that electrons in a piece of copper can move about freely, and so make this metal a *conductor*? To answer this question we have to see what happens when copper atoms form solid copper. Each of the electrons in an atom comes close to an electron in a similar state in the neighboring atoms. But the Pauli exclusion principle says that they can then no longer be in exactly the same state. The energies change, and what was a single level in the isolated atom becomes a *band* of energies in the metal.



For electrons in one atom to interact with electrons in another, they have to get so close that their wave functions overlap. There is just one electron in the outermost shell of a copper atom. As atoms come together, the outermost electrons are the ones whose wave functions overlap the most. The band that they form is broader than those formed by the inner electrons. There is another difference: all of the levels in the bands formed by the inner electrons are *filled*, i.e., occupied by electrons. That is not so for the outermost band. In it is one electron from each atom. But there is room for two electrons with the same energy in each atom, one in each spin state, i.e., one with each value of m_s, namely $+\frac{1}{2}$ and $-\frac{1}{2}$. As a result the outermost band in the copper metal is only half-filled.

That means that the electrons in that band can easily be given more energy. There are other levels in the same band that they can occupy. That happens when they are accelerated by an electric field. That's why this band is called the *conduction band*. The next lower band, on the other hand, is *full*. It has no empty states. Its electrons are unable to accept small amounts of energy.

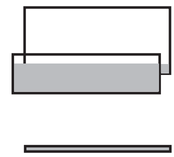
If, on the other hand, there were two electrons from each atom in the uppermost band, then all the available energy levels would be occupied by electrons, and the Pauli exclusion principle would make it impossible for any one electron to go to another energy level in that band. Each electron would be trapped, unable to accept more energy and so be able to move away, and the material would not be able to conduct. It would no longer be a metal, but would, instead, be an *insulator*.



In the figure the shaded parts represent closely spaced energy levels that are occupied by electrons. In the conduction band of the metal there are empty levels, so that the electrons can accept small amounts of energy and occupy them. In the insulator, on the other hand, the levels in the corresponding band are completely occupied, so that none of them can accept small amounts of energy. Only an energy sufficient to move the electron to the next higher band can be absorbed.

It's a good story, but we have a problem: if two electrons from each atom cause the band to be filled so that the material is an insulator, why are the elements of the second column of the periodic table, with two electrons in their outermost shell, magnesium, calcium, barium, etc., not insulators? They are metals, although not really *good* metals. Their metallic properties of electric and thermal conduction and mechanical strength are relatively weak.

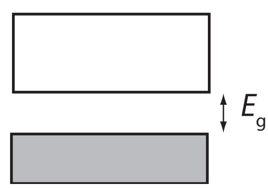
It turns out that what we said is still correct, but in these elements the uppermost energy band, the one that we would expect to be filled, overlaps with the next empty one, and electrons spill



over into it. There are now adjacent empty levels and the electrons are no longer locked into one band. They are free to accept energy, and so to move. We have to wait for the fourth column of the periodic table to get to nonmetals. Here new adventures are waiting.

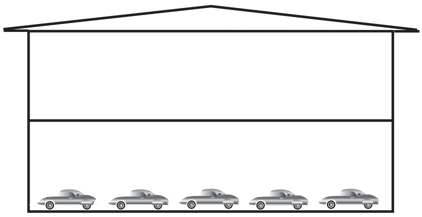
Let's look at the insulator again. The band with the highest-energy electrons is filled. The next one is empty. We are here talking about the *ground state*, the state where electrons occupy the lowest energy levels that they are allowed to be in, one at a time, as limited by the Pauli exclusion principle.

An electron from a filled band can be promoted to a band with empty states if we can give it the required energy. It can be by shining light on it (with just the right energy) or by any other means that accomplishes the same purpose. In any case the minimum energy is the amount that brings the electron from the top of the filled band to the bottom of the next empty band, the energy of the *gap*, E_g, between the two bands.



For some materials the gap is so small that even the internal (*thermal*) energy at room temperature, the energy of the shaking of the atoms of the crystal lattice, is sufficient to get some of the electrons across the gap. In this case, even when the material is an insulator as the temperature approaches absolute zero, there will be some electrons in the *conduction band*, the otherwise empty band across the gap to which the thermal energy can promote them. This is what happens in germanium and silicon, and this is why these elements are *semiconductors*.

Think of a two-floor garage, with the whole floor divided into spaces, each just big enough for one car. If all spaces on the first floor are filled,



none of the cars there can move. Now let one car from the first floor be taken to the empty second floor. There it can move around at will. But in addition something else has happened: the first floor is no longer full. If forces are applied to all cars, to the left, a car to the right of the now empty spot can move to its left. A new empty space is created, and a car to its right can move one step over. The two-floor garage model has been adapted from *The Junction Transistor* by Morgan Sparks in the July 1952 issue of *Scientific American*, where it is attributed to William Shockley.

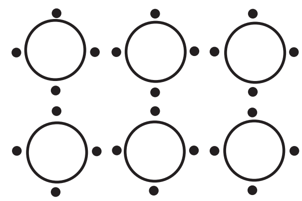
We can also take a new and marvelously fruitful point of view: focus on the empty space, the *hole* in the sea of cars. As cars, one by one, move to the left, the empty space moves to the right. Go back to the electrons. Put on an electric field to the right. Each electron experiences a force to the left, but in a filled band none can move. Take just one electron to the next, the conduction band, and it can contribute to the current, pushed by the field. But the hole left behind in the lower band also responds to the field. It moves in the opposite direction to that of the electron, in the direction of the force on a positive charge.

It is only electrons that move. There are no moving positive charges. But the hole in the electron sea moves in the direction of the electric field *as if* it had a positive charge, and is sometimes said to *have* a positive charge. When an electron is "promoted" to the conduction band there is a double reward: the electron is now able to move, and so is the hole that it leaves behind.

15.2 Tiny changes and vast consequences: impurities in semiconductors

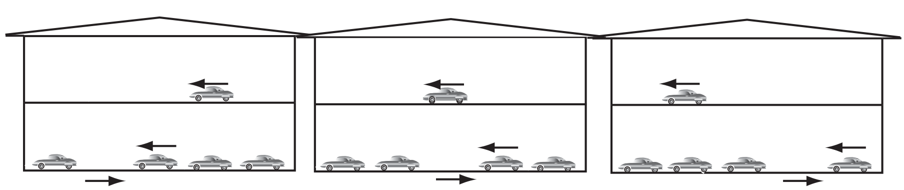
If that were the whole story of semiconductors it would already be great. But there is an even easier way to affect the conduction properties, and this is the one that has made the creation of today's electronic devices possible.

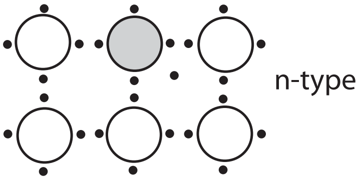
Here is a representation that shows the individual atoms and their outer electrons.

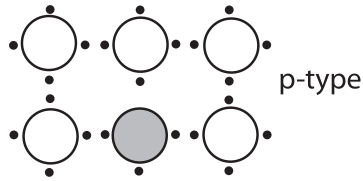


Put in some impurities. Silicon is in the fourth column of the periodic table, element number 14. At absolute zero all its electrons are in filled bands. Each atom contributes four electrons to the highest band. These are the electrons that are responsible for the force between atoms. Each of them forms a bond with one of the electrons in a neighboring atom. The two electrons of the bond are shared by the two atoms in what is called a *covalent* bond.

Now exchange one of the silicon atoms for an atom from the fifth column, such as phosphorus ($Z = 15$), with one more electron in its atom. Four of the electrons take the place of the silicon atom's four outer electrons in holding on to their neighbors with covalent bonds. But what happens to the fifth one? It isn't needed. It stays near "its" atom, but because it is superfluous, because it does not form part of a bond, it is very loosely attached, and can be removed with just a small amount of energy. Once liberated, it is a free electron, free to move around. It is now in the conduction band. The amount of energy needed to get it there is much smaller







than that which would be necessary to pull up an electron across the gap. It is so small that the internal energy at room temperature is sufficient to get a significant number of the electrons there and so to contribute to the conduction process.

p-type, n-type

There is still another wrinkle to add to the versatility of semiconductors. Let the impurity atom be in the third column, like boron, with three outer electrons. The three electrons can take the place of only three of the silicon atom's four bonding electrons. The atom needs a fourth, and is ready to snare one wherever it can find one. It doesn't take much energy to steal one from a neighbor. But the neighbor is now unfulfilled, and it, in turn, pulls one away from some other neighbor. The missing place, *the hole*, travels backward, as before. The material is now called a *p-type semiconductor*, because of the holes that experience forces in the electric field as if they were positively charged. In contrast, the material with extra negatively charged electrons is called an *n-type semiconductor*.

We can use the two-floor garage model to show what happens in an n-type semiconductor: a car is moved to the second floor while the first floor remains filled, that is, an electron is promoted to the conduction band while the next lower band remains full. Similarly, in a p-type semiconductor an empty space appears on the first floor, without any other changes, that is, a hole is formed in the highest band, making conduction possible there, without any electrons in the conduction band.

Note that we have used three different representations: the energy level diagram that shows the energy bands, the two-level garage, and the lattice of individual atoms and electrons.

EXAMPLE 1

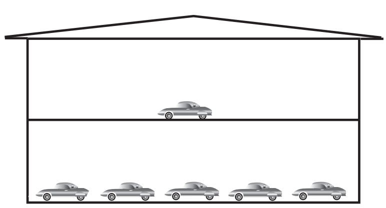
Rank the following in order of their resistivity at room temperature: (a) copper, (b) pure silicon, (c) silicon with 0.01% phosphorus.

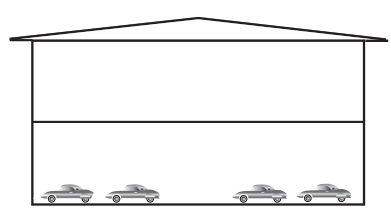
Ans.:
The resisitivity depends primarily on the number of free electrons (the electrons in the conduction band) in a given volume of material.

(a) In copper one electron from each atom is a free electron.

(b) At absolute zero pure silicon has no free electrons and is an insulator. At higher temperatures some of the electrons have enough thermal energy to transcend the energy gap of 1.1 eV so that they are in the conduction band. Since the gap energy is much larger than the thermal energy at room temperature ($kT = 0.025\,\text{eV}$), only a very small number of electrons is in the conduction band.

(c) The atomic number of phosphorus is greater than that of silicon by one. Each phosphorus atom therefore has one more electron than an atom of silicon. The phosphorus atom takes the place of a silicon atom in the crystal lattice. It takes very little energy to remove the extra electron so that it becomes a free electron. Since only 0.01% of the atoms provides free electrons, the resistivity is much greater than that of copper, but much less than that of pure silicon, where





there are no impurities to provide free electrons, and the only free electrons are those that are in the conduction band because of their thermal energy.

The order of increasing resistivity is therefore (a), (c), (b).

EXAMPLE 2

How are the resistivity of a material and the resistance of a piece of wire related to the number of charges that are free to move?

Ans.:
Under the heading *Motion of a charged object in a magnetic field* on page 205 we looked at a wire in which particles with a total charge Q are moving, each with a velocity v parallel to the wire. The current in the wire is I and its length is L. We saw that $IL = Qv$.

If the charge Q consists of lots of charges, each with magnitude e, and their number is n per unit volume, then the charge per unit volume is ne. This is called the *charge density*, for which the SI unit is C/m^3. The total charge in the volume LA is $neLA$.

We see that $IL = neLAv$, which we can rewrite as $\frac{I}{A} = nev$. If we define the current density J to be $J = \frac{I}{A}$, then $J = nev$. (This is an important relation that relates the macroscopic quantitiy J, which describes the current, to the microscopic quantities n, e, and v, which describe the number per unit volume, the charge, and the average speed of the free electrons.)

Since $R = \frac{\Delta V}{I}$, and the resistivity, ρ, is given by $R = \rho\frac{L}{A}$, we see that both are inversely proportional to the charge density, ne.

15.3 The transistor and the information revolution

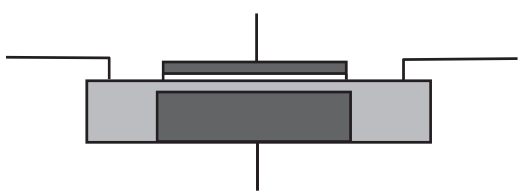
The history of the 1940s is that of war and suffering. The development of the nuclear reactions of fission and fusion for explosives, and later for the utilization of energy, is part of that history. So is the invention of radar. The vast human and technological consequences of these scientific developments tend to overshadow the fact that this was also the time of the *semiconductor revolution*, the beginning of what is sometimes called the *silicon age*.

It is the confluence of two tidal waves of history that led to this revolution. One is based on pure thought, the other on a very down-to-earth practical development, both at the central core of modern physics. One is quantum mechanics, the science that describes atoms and electrons. The other is materials science, the descendant of the metallurgy of the ancients, which characterized the bronze age and the iron age. It made possible the preparation of semiconducting elements in crystalline form with until then unimaginable regularity and chemical purity. The knowledge of one part in 10^8 that was achieved is analogous to knowing the number and position of everyone in the United States to ± 2 individuals.

This is history outside the mainstream of popular culture, too often ignored or misunderstood, but with the most profound consequences for that culture. It is the history that paved the way for how we communicate today. Computers and the internet, cell phones, and portable music players could not exist without it.

The transistor

In semiconductors the transfer of a small fraction of the electrons from one band to another can cause large changes in the conduction properties, and hence in an electric current. Such changes can be produced by the addition of minute amounts of impurities. They can also be brought about by electric fields, which move electrons in and out of narrow current-carrying channels.



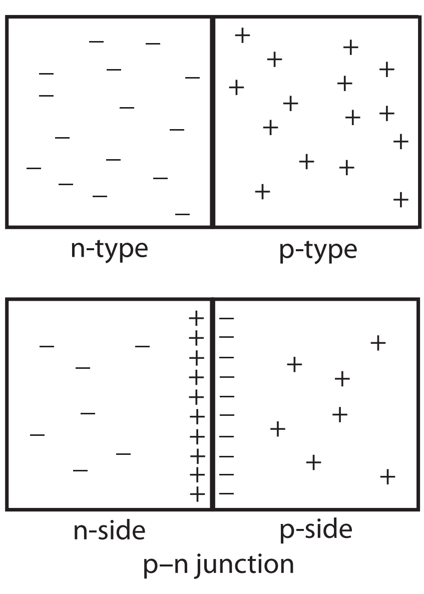
This is what happens in a *field effect transistor*. This was not the first or even the second type of transistor to be made, but it is the one that is simplest to describe, and it is the one in most widespread use today. The concept was patented in 1926 by Julius Lilienfeld, but at that time semiconductor technology was not even in its infancy, and no usable devices were constructed. The idea was reinvented much later, and the first field effect transistors were made at Bell Laboratories in about 1960. Even then it took another 15 years for them to be in common use.

In a field effect transistor the electric field is at right angles to the current in a conducting channel of either an n-type or a p-type semiconductor. The field causes electrons to be pulled in or out of that channel and can even change it from n-type to p-type or the other way around. As a result the current in the channel can change by a large amount.

Since a small change in the field can cause a large change in the current, the transistor can act as an amplifier. It can control the current and switch it on or off. A combination of two transistors in parallel can be used as a *bistable* combination, with the current in either one or the other. This is the fundamental unit of a calculator or computer, where the two possible states are usually referred to as "1" and "0." A great number of such units combine to give us the logical and computational power of today's devices.

The p-n junction

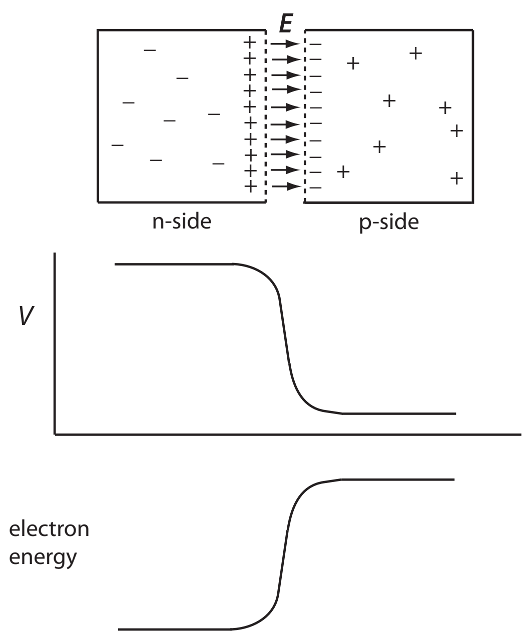
Take a piece of n-type semiconductor (with its extra electrons) and put it together with a piece of p-type material (with its extra holes). Not with glue, which would mess it up, but with the two types of impurity *grown-in*, each on one side of the border region, which is now called a p–n junction.



On the n-type side there are electrons moving about freely, aimlessly. The same is true for the holes on the p-type side. At the junction the electrons on their side are close to the holes on the other side. A hole is a place where an electron is missing, and a passing electron can drift across the border and "fall in."

When electrons do that they accumulate along the junction on the p-side where they have been captured into holes. There are then two layers of charge, negative on the p-side and positive on the n-side, where the electrons have left to cross the boundary. The two charge layers produce a built-in electric field between them. The field builds up until there is so much negative charge on the p-side that it repels further negative charges so strongly that the migration stops.

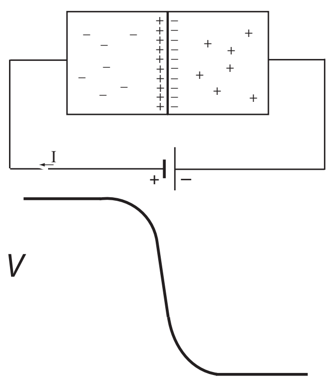
In this equilibrium configuration there is now a barrier that repels electrons as they try to drift to the p-side and holes as they try to enter the n-side. In other words the barrier is such that both kinds of charge carriers are kept "home," the electrons on the n-side and the holes on the p-side.



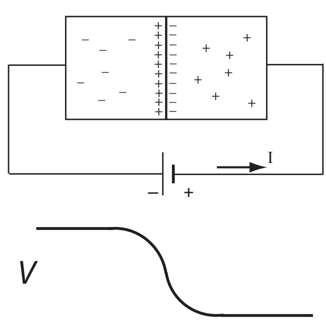
We can also look at the electric potential. The n-side, with its extra positive charge, is at a higher potential than the p-side. The field, as always, is in the direction from the higher potential to the lower potential.

A third representation is that of the energy of the electrons. Since the electrons are negatively charged, their potential energy increases as the electric potential decreases. While they "fall up" on the graph of the electric potential, they fall down on the energy diagram. The figure shows all three: the built-in field in the junction region, the variation of the electric potential, and the variation of the electron energy.

Suppose that we connect the two sides of the p–n junction to a battery, with the n-side connected to the positive terminal and the p-side to the negative termiinal. It is now even harder for the holes to get to the n-side and the electrons to get to the p-side. The height of the barrier has increased and any flow of charge is inhibited.



barrier increased



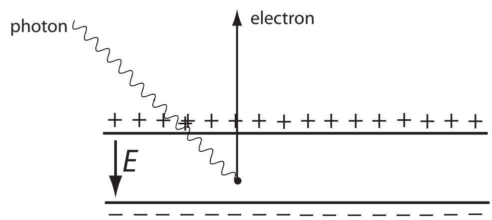
barrier decreased

The result is quite different if we connect the positive terminal of the battery to the p-side of the junction and the negative terminal to the n-side. The barrier is now reduced. In the equilibrium situation it just barely held the charges in their home territory. When the height of the barrier is reduced, charges will flow across it as over a waterfall.

We see that charges flow easily in one direction, but not in the other. The p–n junction acts as a rectifier.

The solar cell

Fabricate the junction so that it is near the surface. Shine a beam of light on it. By the photoelectric effect a photon can give its energy to an electron near the junction region, perhaps enough to liberate it and make it a free electron.



If it gains its freedom in the region of the built-in electric field, it will feel a force, and it will be pushed across the border, back to the n-side. If this happens to lots of electrons they will constitute a current, to which the energy of the photons has been transferred. If the photons come from the sun, we have transformed solar energy into electric energy.

The raw material, the silicon, costs next to nothing. The solar energy is free. The manufacturing process, however, and the auxiliary network of contacts and wires, is so expensive that up to this point solar cells are used only to a very small extent, mostly in special applications where cost is not the main factor. Predictions of a turn in the economic balance have not yet been realized, but it is widely thought that this will eventually happen.

The future

We are surrounded, at least in the developed world, by the fruits of electron science. Communications depend on it, and so does a large fraction of modern entertainment. Most recently, and still growing explosively, there is the information industry. Telephone, television, the recording of sound and image, and the computer have transformed our lives and moved the boundaries of the possible.

Some adventurous people call themselves futurists, and try to predict new technologies and the challenges that they will pose. As we look back we see that they have rarely been equal to the task. One of the pioneers of the computer age was asked what he thought computers might be good for. His answer was that they would allow accurate weather prediction. Certainly no one thought that they would displace typewriters, or give rise to message networks.

The development of science and technology has been fanciful and unpredictable. As we look to the future we must expect it to continue to be so. Our powers of prediction, on the other hand, depend on our present knowledge. There are technologies waiting to be exploited, such solar cells and refrigerators based on solid-state science. Our understanding of the structure of matter and the behavior of electrons in solids has led to the development of new materials that promise new applications of magnetism and electricity in ways that we can only dimly foresee. Our knowledge is being applied to more and more complex systems, particularly in biology. We don't know where it will lead us, but the new developments will, no doubt, be informed by the past, even as they lead into the unknown.

Modern technology touches our lives in many ways. It is not always looked on as favoring the common good, and there are strong reactions against its pervasive presence. In some quarters science is looked on with suspicion and with hostility.

Does our educational system prepare us to live in an age in which science plays such an important role? The presence of science and technology in our lives has generally not brought about a corresponding increase in widespread knowledge about these areas. Yet citizens and their representatives are asked to make choices and judgments that channel private and public activities and resources and profoundly affect all of us.

It is hard to know whether, individually and collectively, we will bring wisdom to our choices. Surely, however, a prerequisite is that we know what we are talking about. It is toward the increase in that knowledge that this book is dedicated.

15.4 Summary

When atoms cluster together to form a solid, each electron energy level of the atom becomes a *band* of energy levels in the solid. Each energy level in the band may be occupied by an electron or it may be empty.

If all of the bands of a solid are either completely full or completely empty, the solid is an insulator. If one or more of the energy bands is only partially filled, the material is a metal. A special situation occurs when the bands are fully occupied up to a certain energy at the absolute zero of temperature, but the next band is so close in energy that thermal energies are sufficient to cause electrons to go to it. The material is then a semiconductor.

A semiconductor may have impurities with extra electrons that are easily removed so as to become *conduction electrons*, i.e., electrons that are free to move in a band that was previously empty. Such a material is called an *n-type semiconductor*.

Impurities may also have less electrons so that electrons from the semiconductor can jump to it, leaving a missing electron or *hole* in a band that was previously filled. The hole behaves like a positive electron, and the material is called a *p-type semiconductor*.

In a *field effect transistor* an electric field can be applied to change the number of electrons or holes in a current-carrying channel. A very small amount of energy can then change or control a current with a much larger energy.

A boundary where an n-type semiconductor changes to a p-type semiconductor is called a *p–n junction*. Electrons migrate to the p-side and an electric field is set up at the boundary from the n-side to the p-side. The p–n junction conducts easily in only one direction so that it acts as a rectifier. An electron liberated in the junction region by the photoelectric effect is propelled by the electric field. This is the principle of the solar cell.

Our ability to predict the future course of science and technology is limited. Based on past experience, we can expect new and surprising developments that have profound effects on the way we live.

15.5 Review activities and problems

Guided review

1. (a) Describe the difference between an insulator and a metal in terms of their energy bands.

(b) Describe the difference between an insulator and a semiconductor in terms of their energy bands.

(c) Describe n-type and p-type semiconductors.

2. (a) What is the number of free electrons per m^3 in copper? (You may need its atomic mass, 63.5, and its density, 8.9×10^3 kg/m^3.)

(b) What is the average or drift velocity of the electrons in a wire with cross section 1 mm^2 carrying a current of 10 A?

Problems and reasoning skill building

1. Use the answer to *Guided review* question 2 to answer the following questions:

(a) How many states are there in the conduction band of 1 g of copper?

(b) What fraction of these states are occupied by electrons?

2. (a) Sulfur is an insulator. What does that tell you about its energy bands?

(b) Calcium is a metal. What does that tell you about its energy bands?

(c) Can you tell that calcium is a metal from its position in the periodic table of elements? What additional information do you need?

3. (a) In the ground state of silicon the electrons with the highest energy are 3p electrons, i.e., electrons with $n = 3$ and $\ell = 1$. Rank all of the electrons in the ground state of silicon in the order of their energy, from the lowest to the highest.

(b) For which of the electrons does the orbit (the dominant part of the wave function) have the largest diameter? The smallest?

4. (a) What is the number, n, of free electrons per m^3 in silicon with 0.01% of phosphorus? (Assume that all of the extra electrons provided by the phosphorus are free. The atomic mass of silicon is 28.1 and its density is 2.3×10^3 kg/m^3.)

(b) Assume that the drift velocity is the same as in the copper wire of *Guided review* question 2. What is then the resistivity of the silicon with the phosphorus impurity?

5. A free electron appears in the junction region of a p–n junction. Toward which side will it experience a force? Explain this with the help of a diagram.

Synthesis problems and projects

1. For silicon with phosphorus impurities we have shown two representations for the electron behavior: the lattice of atoms and electrons and the two-level garage. Now show the upper energy bands for silicon. In them show the level for the liberated conduction electron provided by the phosphorus impurity atoms.

Then decide where the level is before the electron is free. (Do this from your knowledge that the electron is much more weakly bound than the silicon electrons.)

2. (a) A hole is created in pure silicon when an electron "jumps across the gap." Show this transition on an energy level diagram. Show the level of the electron and of the hole that it leaves behind.

(b) It takes much less energy to create a hole in silicon with a small amount of boron impurity. Show the level of the hole that is created in this case and an arrow showing the transition as the hole is created.