14 Semiconductors (extracted from the Richard Feyman lectures)

14–1 Electrons and holes in semiconductors

There are large numbers of semiconductors known, but we'll concentrate on those which now have the greatest technical application. They are also the ones that are best understood, and in understanding them we will obtain a degree of understanding of many of the others. The semiconductor substances in most common use today are silicon and germanium. These elements crystallize in the diamond lattice, a kind of cubic structure in which the atoms have tetrahedral bonding with their four nearest neighbors. They are insulators at very low temperatures—near absolute zero—although they do conduct electricity somewhat at room temperature. They are not metals; they are called *semiconductors*.

If we somehow put an extra electron into a crystal of silicon or germanium which is at a low temperature, we will have just the situation we described in the last chapter. The electron will be able to wander around in the crystal jumping from one atomic site to the next. Actually, we have looked only at the behavior of electrons in a rectangular lattice, and the equations would be somewhat different for the real lattice of silicon or germanium. All of the essential points are, however, illustrated by the results for the rectangular lattice.

These electrons can have energies only in a certain energy band—called the *conduction band*. Within this band the energy is related to the wave-number k of the probability amplitude C by

$$E = E_0 - 2A_x \cos k_x a - 2A_y \cos k_y b - 2A_z \cos k_z c. \qquad (14.1)$$

The A's are the amplitudes for jumping in the x-, y-, and z-directions, and a, b, and c are the lattice spacings in these directions.

For energies near the bottom of the band, we can approximate Eq. ([14.1](#)) by

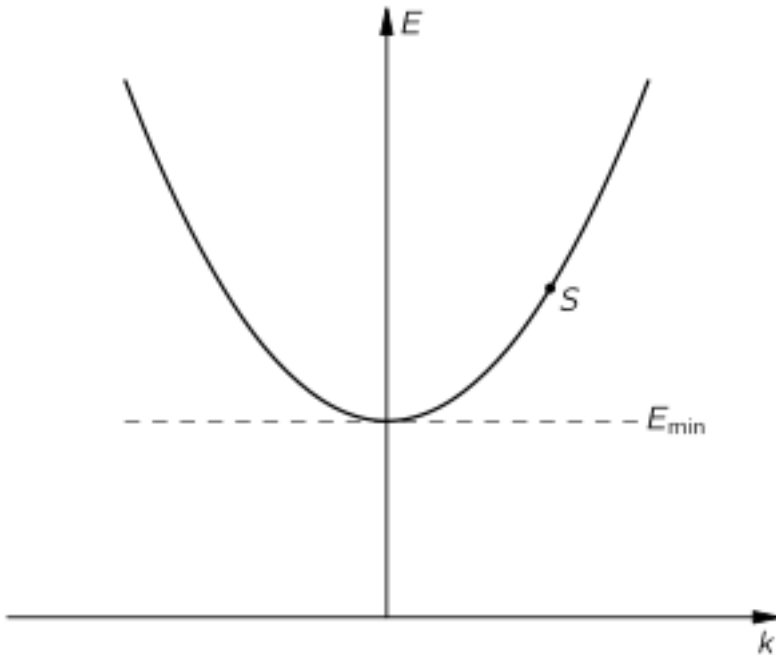$$E = E_{min} + A_x a^2 k_x^2 + A_y b^2 k_y^2 + A_z c^2 k_z^2 \qquad (14.2)$$

Fig. 14–1.The energy diagram for an electron in an insulating crystal.

If we think of electron motion in some particular direction, so that the components of k are always in the same ratio, the energy is a quadratic function of the wave number—and as we have seen of the momentum of the electron. We can write

$$E = E_{min} + \alpha k^2, \qquad (14.3)$$

where α is some constant, and we can make a graph of E versus k as in Fig. 14–1. We'll call such a graph an "energy diagram." An electron in a particular state of energy and momentum can be indicated by a point such as S in the figure.

We can have a similar situation if we *remove* an electron from a neutral insulator. Then, an electron can jump over from a nearby atom and fill the "hole," but leaving another "hole" at the atom it started from. We can describe this behavior by writing an amplitude to find the *hole* at any particular atom, and by saying that the *hole* can jump from one atom to the next. (Clearly, the amplitudes A that the hole jumps from atom a to atom b is just the same as the amplitude that an electron on atom b jumps into the hole at atom a.) The mathematics is just the same for the *hole* as it was for the extra electron, and we get again that the energy of the hole is related to its wave number by an equation just like Eq. (14.1) or (14.2), except, of course, with different numerical values for the amplitudes A_x, A_y, and A_z. The hole has an energy related to the wave number of its probability amplitudes. Its energy lies in a restricted band, and near the bottom of the band its energy varies quadratically with the wave number—or momentum—just as in Fig. 14–1.

We also find that *the hole also behaves like a classical particle* with a certain effective mass—except that in noncubic crystals the mass depends on the direction of motion. So the hole behaves like a *positive particle* moving through the crystal. The charge of the hole-particle is positive, because it is located at the site of a missing electron; and when it moves in one direction there are

actually electrons moving in the opposite direction.

If we put several electrons into a neutral crystal, they will move around much like the atoms of a low-pressure gas. If there are not too many, their interactions will not be very important. If we then put an electric field across the crystal, the electrons will start to move and an electric current will flow. Eventually they would all be drawn to one edge of the crystal, and, if there is a metal electrode there, they would be collected, leaving the crystal neutral.

Similarly we could put many holes into a crystal. They would roam around at random unless there is an electric field. With a field they would flow toward the negative terminal, and would be "collected"—what actually happens is that they are neutralized by electrons from the metal terminal.

One can also have both holes and electrons together. If there are not too many, they will all go their way independently. With an electric field, they will all contribute to the current. For obvious reasons, electrons are called the *negative carriers* and the holes are called the *positive carriers*.

We have so far considered that electrons are put into the crystal from the outside, or are removed to make a hole. It is also possible to "create" an electron-hole pair by taking a bound electron away from one neutral atom and putting it some distance away in the same crystal. We then have a free electron and a free hole, and the two can move about as we have described.
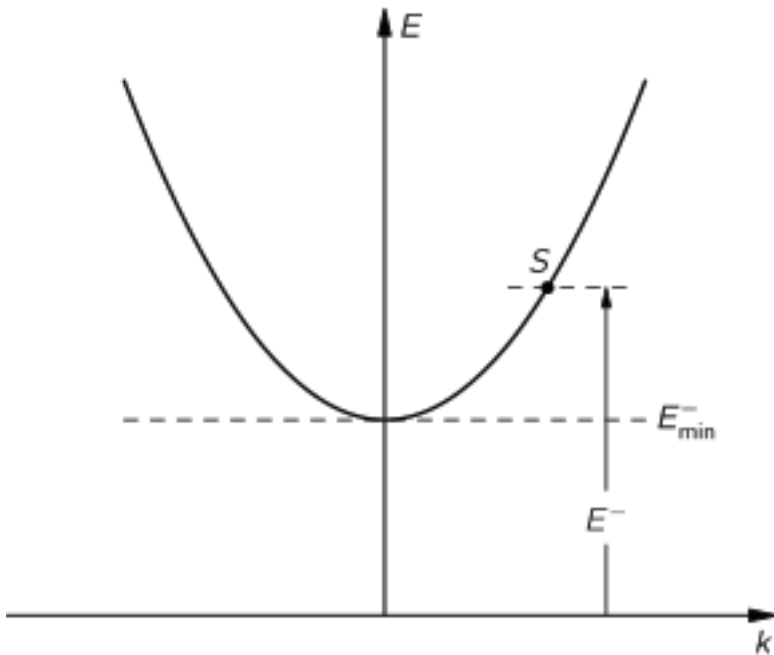

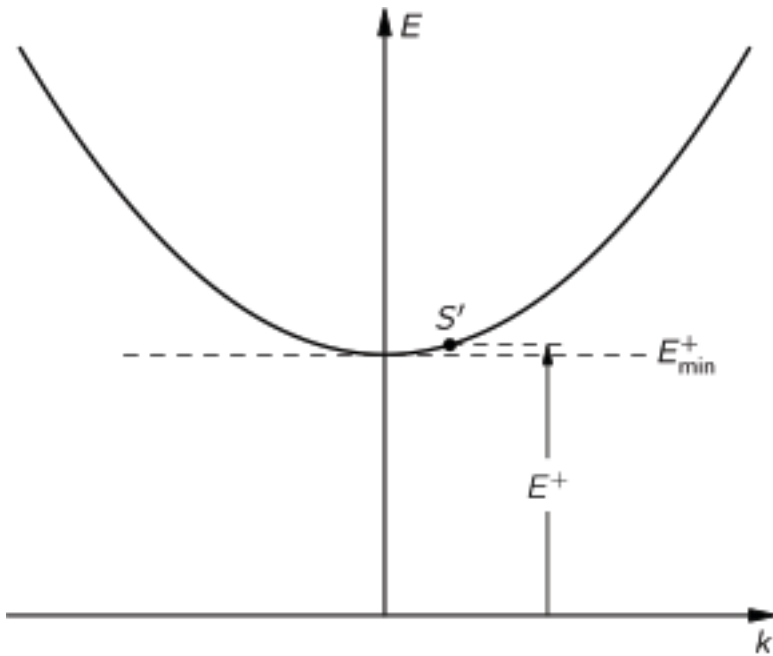Fig. 14–2.The energy E^- is required to "create" a free electron.

Fig. 14–3.The energy E^+ is required to "create" a hole in the state S'.

The energy required to put an electron *into* a state S—we say to "create" the state S—is the energy E^- shown in Fig. 14–2. It is some energy above $E^-_{\min}$. The energy required to "create" a hole in some state S' is the energy E^+ of Fig. 14–3, which is some energy greater than $E^+_{\min}$. Now if we create a pair in the states S and S', the energy required is just $E^- + E^+$.

The creation of pairs is a common process (as we will see later), so many people like to put Fig. 14–2 and Fig. 14–3 together on the same graph—with the *hole* energy plotted *downward*, although it is, of course a *positive* energy. We have combined our two graphs in this way in Fig. 14–4. The advantage of such a graph is that the energy $E_{\text{pair}} = E^- + E^+$ required to create a pair with the electron in S and the hole in S' is just the vertical distance between S and S' as shown in Fig. 14–4. The minimum energy required to create a pair is called the "gap" energy and is equal to $E^-_{\min} + E^+_{\min}$.
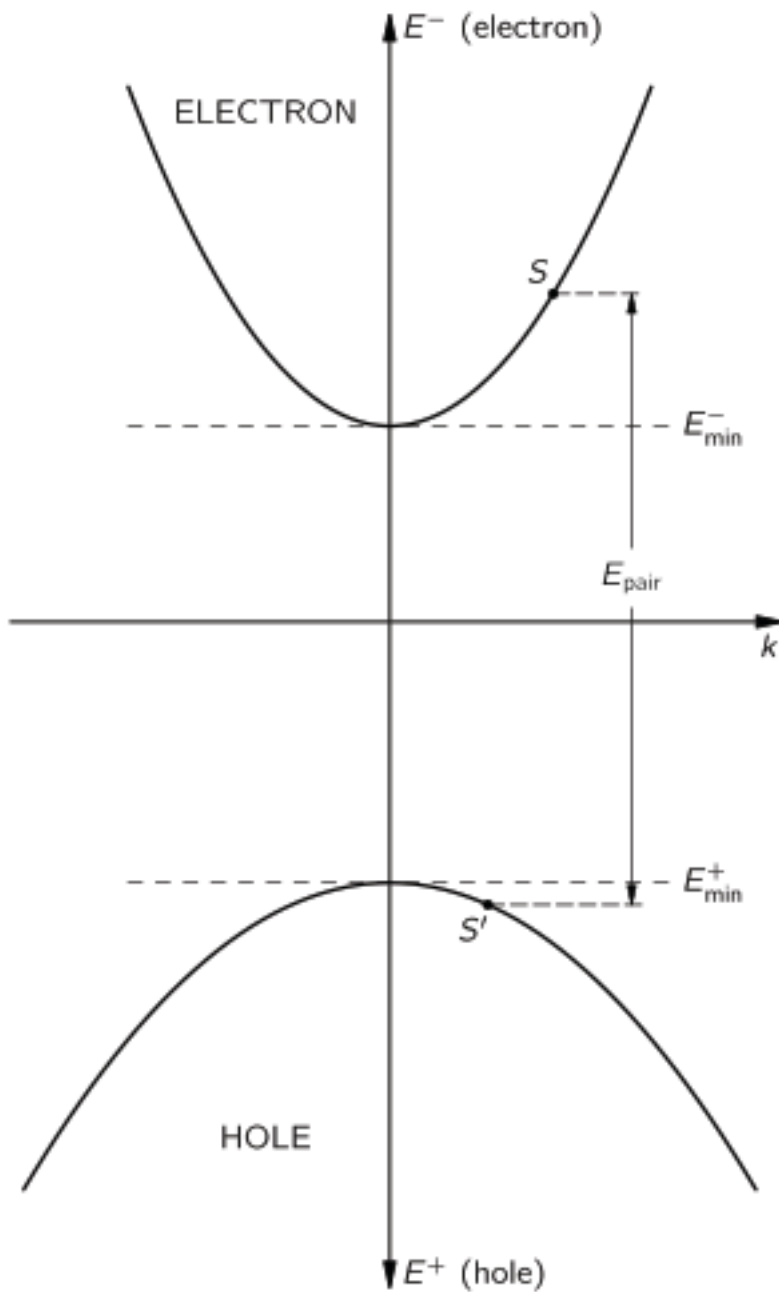
Fig. 14–4.Energy diagrams for an electron and a hole drawn together.

Sometimes you will see a simpler diagram called an energy level diagram which is drawn when people are not interested in the k variable. Such a diagram—shown in Fig. 14–5—just shows the possible energies for the electrons and holes.[1]
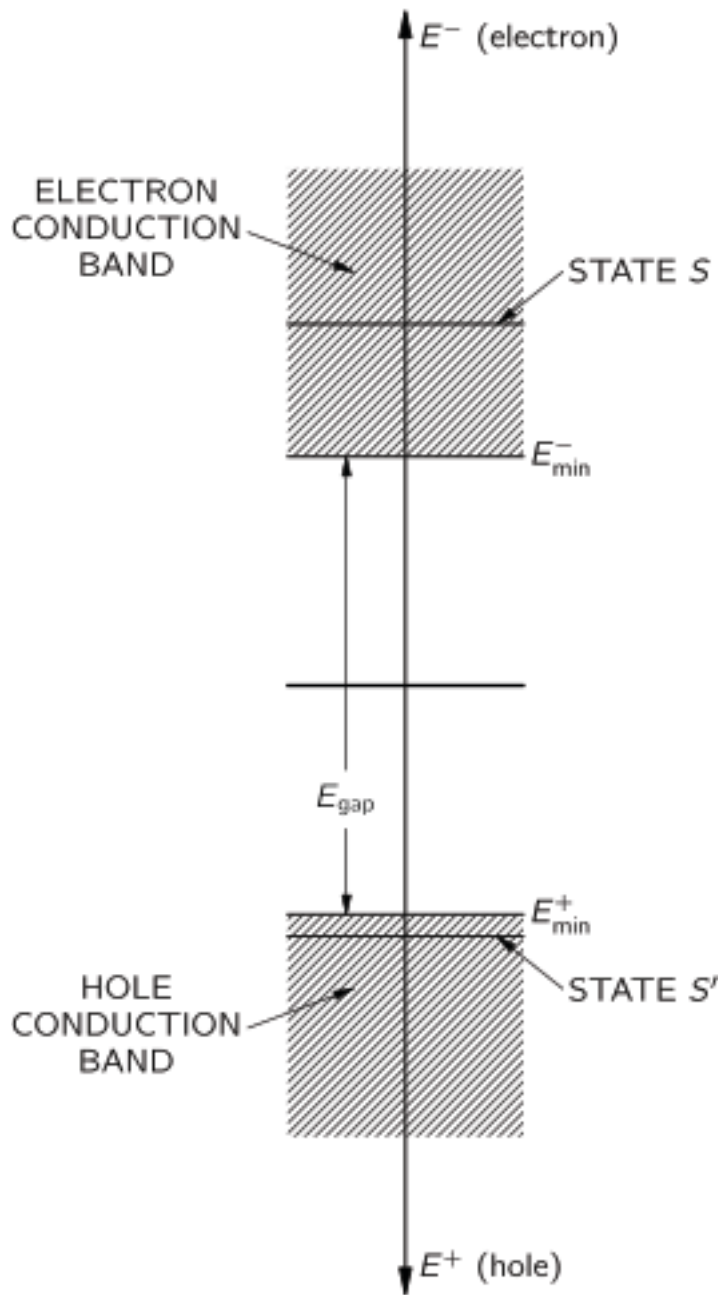
Fig. 14–5.Energy level diagram for electrons and holes.

How can electron-hole pairs be created? At any
finite temperature there is still another mechanism by which electron-hole pairs can be created. The pair
energy can be provided from the thermal energy of the crystal. The thermal vibrations of the crystal can
transfer their energy to a pair—giving rise to "spontaneous" creation.

The probability per unit time that the energy as large as the gap energy E_{gap} will be concentrated at one
atomic site is proportional to $e^{-E_{gap}/\kappa T}$, where T is the temperature and κ is Boltzmann's constant.
Near absolute zero there is no appreciable probability, but as the temperature rises
there is an increasing probability of producing such pairs. At any finite temperature the production
should continue forever at a constant rate giving more and more negative and positive carriers. Of course
that does not happen because after a while the electrons and holes accidentally find each other—the
electron drops into the hole and the excess energy is given to the lattice. We say that the electron and
hole "annihilate." There is a certain probability per second that a hole meets an electron and the two
things annihilate each other.

If the number of electrons per unit volume is N_n (n for negative carriers) and the density of positive
carriers is N_p, the chance per unit time that an electron and a hole will find each other and annihilate is
proportional to the product $N_n N_p$. In equilibrium this rate must equal the rate that pairs are created.
You see that in equilibrium the product of N_n and N_p should be given by some constant times the
Boltzmann factor:

$$N_n N_p = \text{const } e^{-E_{gap}/\kappa T}. \tag{14.4}$$

When we say constant, we mean nearly constant. A more complete theory—which includes more details
about how holes and electrons "find" each other—shows that the "constant" is slightly dependent upon
temperature, but the major dependence on temperature is in the exponential.

Let's consider, as an example, a pure material which is originally neutral. At a finite temperature you
would expect the number of positive and negative carriers to be equal, $N_n = N_p$. Then each of them
should vary with temperature as $e^{-E_{gap}/2\kappa T}$. The variation of many of the properties of a
semiconductor—the conductivity for example—is mainly determined by the exponential factor because
all the other factors vary much more slowly with temperature. The gap energy for germanium is about
0.72 eV and for silicon 1.1 eV.

At room temperature κT is about $1/40$ of an electron volt. At these temperatures there are enough
holes and electrons to give a significant conductivity, while at, say, $30°K$—one-tenth of room
temperature—the conductivity is imperceptible. The gap energy of diamond is 6 or 7 eV and diamond is
a good insulator at room temperature.

14–2 Impure semiconductors

There is still another way to put
electrons into the conduction band of a crystal. Suppose we imagine a crystal of
germanium in which one of the germanium atoms is replaced by an arsenic atom. The germanium atoms
have a valence of 4 and the crystal structure is controlled by the four valence electrons. Arsenic, on the
other hand, has a valence of 5. It turns out that a single arsenic atom can sit in the germanium lattice
(because it has approximately the correct size), but in doing so it must act as a valence 4 atom—using
four of its valence electrons to form the crystal bonds and having one electron left over. This extra
electron is very loosely attached—the binding energy is only about 1/100 of an electron volt. At room
temperature the electron easily picks up that much energy from the thermal energy of the crystal, and
then takes off on its own—moving about in the lattice as a free electron. An impurity atom such as the
arsenic is called a *donor site* because it can give up a negative carrier to the crystal. If a crystal of
germanium is grown from a melt to which a very small amount of arsenic has been added, the arsenic
donor sites will be distributed throughout the crystal and the crystal will have a certain density of
negative carriers built in.

You might think that these carriers would get swept away as soon as any small electric field was put
across the crystal. This will not happen, however, because the arsenic atoms in the body of the crystal
each have a positive charge. If the body of the crystal is to remain neutral, the average density of
negative carrier electrons must be equal to the density of donor sites. If you put two electrodes on the
edges of such a crystal and connect them to a battery, a current will flow; but as the carrier electrons are
swept out at one end, new conduction electrons must be introduced from the electrode on the other end
so that the average density of conduction electrons is left very nearly equal to the density of donor sites.

Since the donor sites are positively charged, there will be some tendency for them to capture some of the
conduction electrons as they diffuse around inside the crystal. A donor site can, therefore, act as a trap
such as those we discussed in the last section. But if the trapping energy is sufficiently small—as it is for
arsenic—the number of carriers which are trapped at any one time is a small fraction of the total. For a
complete understanding of the behavior of semiconductors one must take into account this trapping. For
the rest of our discussion, however, we will assume that the trapping energy is sufficiently low and the
temperature is sufficiently high, that all of the donor sites have given up their electrons. This is, of
course, just an approximation.

It is also possible to build into a germanium crystal some impurity atom whose valence is 3, such as
aluminum. The aluminum atom tries to act as a valence 4 object by stealing an extra electron. It can
steal an electron from some nearby germanium atom and end up as a negatively charged atom with an
effective valence of 4. Of course, when it steals the electron from a germanium atom, it leaves a hole
there; and this hole can wander around in the crystal as a positive carrier. An impurity atom which can
produce a hole in this way is called an *acceptor* because it "accepts" an electron. If a germanium or a
silicon crystal is grown from a melt to which a small amount of aluminum impurity has been added, the
crystal will have built-in a certain density of holes which can act as positive carriers.

When a donor or an acceptor impurity is added to a semiconductor, we say that the material has been
"doped."

When a germanium crystal with some built-in donor impurities is at room temperature, some conduction
electrons are contributed by the thermally induced electron-hole pair creation as well as by the donor
sites. The electrons from both sources are, naturally, equivalent, and it is the total number N_n which

comes into play in the statistical processes that lead to equilibrium. If the temperature is not too low, the number of negative carriers contributed by the donor impurity atoms is roughly equal to the number of impurity atoms present. In equilibrium Eq. (14.4) must still be valid; at a given temperature the product $N_n N_p$ is determined. This means that if we add some donor impurity which increases N_n, the number N_p of positive carriers will have to decrease by such an amount that $N_p N_n$ is unchanged. If the impurity concentration is high enough, the number N_n of negative carriers is determined by the number of donor sites and is nearly independent of temperature—all of the variation in the exponential factor is supplied by N_p, even though it is much less than N_n. An otherwise pure crystal with a small concentration of donor impurity will have a majority of negative carriers; such a material is called an "n-type" semiconductor.

If an acceptor-type impurity is added to the crystal lattice, some of the new holes will drift around and annihilate some of the free electrons produced by thermal fluctuation. This process will go on until Eq. (14.4) is satisfied. Under equilibrium conditions the number of positive carriers will be increased and the number of negative carriers will be decreased, leaving the product a constant. A material with an excess of positive carriers is called a "p-type" semiconductor.

If we put two electrodes on a piece of semiconductor crystal and connect them to a source of potential difference, there will be an electric field inside the crystal. The electric field will cause the positive and the negative carriers to move, and an electric current will flow. Let's consider first what will happen in an n-type material in which there is a large majority of negative carriers. For such material we can disregard the holes; they will contribute very little to the current because there are so few of them. In an ideal crystal the carriers would move across without any impediment. In a real crystal at a finite temperature, however,—especially in a crystal with some impurities—the electrons do not move completely freely. They are continually making collisions which knock them out of their original trajectories, that is, changing their momentum. These collisions are just exactly the scatterings we talked about in the last chapter and occur at any irregularity in the crystal lattice. In an n-type material the main causes of scattering are the very donor sites that are producing the carriers. Since the conduction electrons have a very slightly different energy at the donor sites, the probability waves are scattered from that point. Even in a perfectly pure crystal, however, there are (at any finite temperature) irregularities in the lattice due to thermal vibrations. From the classical point of view we can say that the atoms aren't lined up exactly on a regular lattice, but are, at any instant, slightly out of place due to their thermal vibrations. We would like now to make an estimate of the electrical conductivity of such a material.

When an electric field is applied to an n-type semiconductor, each negative carrier will be accelerated in this field, picking up velocity until it is scattered from one of the donor sites. This means that the carriers which are ordinarily moving about in a random fashion with their thermal energies will pick up an average drift velocity along the lines of the electric field and give rise to a current through the crystal. The drift velocity is in general rather small compared with the typical thermal velocities so that we can estimate the current by assuming that the average time that the carrier travels between scatterings is a constant. Let's say that the negative carrier has an effective electric charge q_n. In an electric field E, the force on the carrier will be $q_n E$.

The average drift velocity is given by $F\tau/m$, where F is the force on the charge, τ is

the mean free time between collisions, and m is the mass. We should use the effective mass we calculated in the last chapter but since we want to make a rough calculation we will suppose that this effective mass is the same in all directions. Here we will call it m_n. With this approximation the average drift velocity will be

$$v_{drift} = \frac{q_n E \tau_n}{m_n}.$$

(14.5)

Knowing the drift velocity we can find the current. Electric current density j is just the number of carriers per unit volume, N_n, multiplied by the average drift velocity, and by the charge on each carrier. The current density is therefore

$$j = N_n v_{drift} q_n = \frac{N_n q_n^2 \tau_n}{m_n} E.$$

(14.6)

We see that the current density is proportional to the electric field; such a semiconductor material obeys Ohm's law. The coefficient of proportionality between j and E, the conductivity σ, is

$$\sigma = \frac{N_n q_n^2 \tau_n}{m_n}.$$

(14.7)

For an n-type material the conductivity is relatively independent of temperature. First, the number of majority carriers N_n is determined primarily by the density of donors in the crystal (so long as the temperature is not so low that too many of the carriers are trapped). Second, the mean time between collisions τ_n is mainly controlled by the density of impurity atoms, which is, of course, independent of the temperature.

We can apply all the same arguments to a p-type material, changing only the values of the parameters which appear in Eq. (14.7). If there are comparable numbers of both negative and positive carriers present at the same time, we must add the contributions from each kind of carrier. The total conductivity will be given by

$$\sigma = \frac{N_n q_n^2 \tau_n}{m_n} + \frac{N_p q_p^2 \tau_p}{m_p}.$$

(14.8)

For very pure materials, N_p and N_n will be nearly equal. They will be smaller than in a doped material, so the conductivity will be less. Also they will vary rapidly with temperature (like $e^{-E_{gap}/2\kappa T}$, as we have seen), so the conductivity may change extremely fast with temperature.
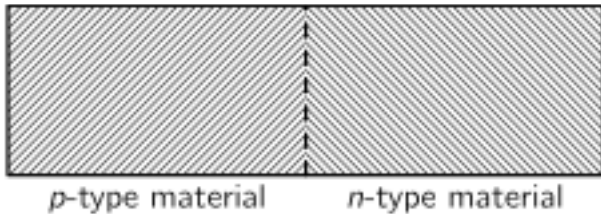
14–4 Semiconductor junctions

Fig. 14–8.A p-n junction.

We would like to discuss now what happens if we take two pieces of germanium or silicon with different internal characteristics—say different kinds or amounts of doping—and put them together to make a "junction." Let's start out with what is called a p-n junction in which we have p-type germanium on one side of the boundary and n-type germanium on the other side of the boundary—as sketched in Fig. 14–8. Actually, it is not practical to put together two separate pieces of crystal and have them in uniform contact on an atomic scale. Instead, junctions are made out of a single crystal which has been modified in the two separate regions. One way is to add some suitable doping impurity to the "melt" after only half of the crystal has grown. Another way is to paint a little of the impurity element on the surface and then heat the crystal causing some impurity atoms to diffuse into the body of the crystal. Junctions made in these ways do not have a sharp boundary, although the boundaries can be made as thin as 10^{-4} centimeters or so. For our discussions we will imagine an ideal situation in which these two regions of the crystal with different properties meeting at a sharp boundary.
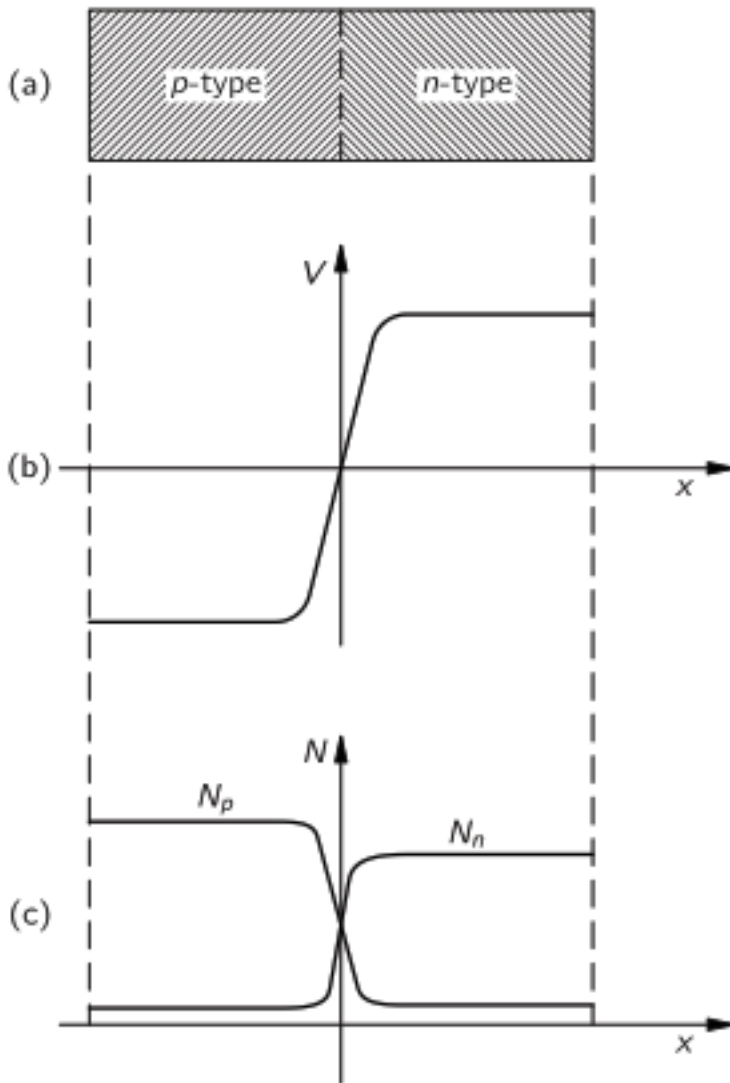
Fig. 14–9.The electric potential and the carrier densities in an unbiased semiconductor junction.

On the n-type side of the p-n junction there are free electrons which can move about, as well as the fixed donor sites which balance the overall electric charge. On the p-type side there are free holes moving about and an equal number of negative acceptor sites keeping the charge balanced. Actually, that describes the situation before we put the two materials in contact. Once they are connected together the situation will change near the boundary. When the electrons in the n-type material arrive at the boundary they will not be reflected back as they would at a free surface, but are able to go right on into the p-type material. Some of the electrons of the n-type material will, therefore, tend to diffuse over into the p-type material where there are fewer electrons. This cannot go on forever because as we lose electrons from the n-side the net positive charge there increases until finally an electric voltage is built up which retards the diffusion of electrons into the p-side. In a similar way, the positive carriers of the p-type material can diffuse across the junction into the n-type material. When they do this they leave behind an excess of negative charge. Under equilibrium conditions the net diffusion current must be zero. This is brought about by the electric fields, which are established in such a way as to draw the positive carriers back toward the p-type material.

The two diffusion processes we have been describing go on simultaneously and, you will notice, both act in the direction which will charge up the n-type material in a positive sense and the p-type material in a negative sense. Because of the finite conductivity of the semiconductor material, the change in

potential from the p-side to the n-side will occur in a relatively narrow region near the boundary; the main body of each block of material will have a uniform potential. Let's imagine an x-axis in a direction perpendicular to the boundary surface. Then the electric potential will vary with x, as shown in Fig. 14–9(b). We have also shown in part (c) of the figure the expected variation of the density N_n of n-carriers and the density N_p of p-carriers. Far away from the junction the carrier densities N_p and N_n should be just the equilibrium density we would expect for individual blocks of materials at the same temperature. (We have drawn the figure for a junction in which the p-type material is more heavily doped than the n-type material.) Because of the potential gradient at the junction, the positive carriers have to climb up a potential hill to get to the n-type side. This means that under equilibrium conditions there can be fewer positive carriers in the n-type material than there are in the p-type material. Remembering the laws of statistical mechanics, we expect that the ratio of p-type carriers on the two sides to be given by the following equation:

$$\frac{N_p(\text{n-side})}{N_p(\text{p-side})} = e^{-q_p V/\kappa T}. \tag{14.10}$$

The product $q_p V$ in the numerator of the exponential is just the energy required to carry a charge of q_p through a potential difference V.

We have a precisely similar equation for the densities of the n-type carriers:

$$\frac{N_n(\text{n-side})}{N_n(\text{p-side})} = e^{-q_n V/\kappa T}. \tag{14.11}$$

If we know the equilibrium densities in each of the two materials, we can use either of the two equations above to determine the potential difference across the junction.

Notice that if Eqs. (14.10) and (14.11) are to give the same value for the potential difference V, the product $N_p N_n$ must be the same for the p-side as for the n-side. (Remember that $q_n = -q_p$.) We have seen earlier, however, that this product depends only on the temperature and the gap energy of the crystal. Provided both sides of the crystal are at the same temperature, the two equations are consistent with the same value of the potential difference.

Since there is a potential difference from one side of the junction to the other, it looks something like a battery. Perhaps if we connect a wire from the n-type side to the p-type side we will get an electrical current. That would be nice because then the current would flow forever without using up any material and we would have an infinite source of energy in violation of the second law of thermodynamics! There is, however, no current if you connect a wire from the p-side to the n-side. And the reason is easy to see. Suppose we imagine first a wire made out of a piece of undoped material. When we connect this wire to the n-type side, we have a junction. There will be a potential difference across this junction. Let's say that it is just one-half the potential difference from the p-type material to the n-type material. When we connect our undoped wire to the p-type side of the junction, there is also a potential difference at this junction—again, one-half the potential drop across the p-n junction. At all the junctions the potential differences adjust themselves so that there is no net current flow in the circuit. Whatever kind of wire you use to connect together the two sides of the n-p junction, you are producing two new junctions, and so long as all the junctions are at the same temperature, the potential jumps at the junctions all compensate each other and no current will flow in the circuit.

We have been assuming that a hole or an
electron produced in a junction region has a good chance of getting into the main body of the crystal
before recombining. The typical time for an electron or a hole to find an opposite partner and annihilate
it is for typical semiconductor materials in the range between 10^{-3} and 10^{-7} seconds. This time is,
incidentally, much longer than the mean free time τ between collisions with scattering sites in the crystal
which we used in the analysis of conductivity. In a typical n-p junction, the time for an electron or hole
formed in the junction region to be swept away into the body of the crystal is generally much shorter
than the recombination time. Most of the pairs will, therefore, contribute to an external current.

14–5Rectification at a semiconductor junction

We would like to show next how it is that a p-n junction can act like a rectifier. If we put a voltage
across the junction, a large current will flow if the polarity is in one direction, but a very small current
will flow if the same voltage is applied in the opposite direction. If an alternating voltage is applied
across the junction, a net current will flow in one direction—the current is "rectified." Let's look again at
what is going on in the equilibrium condition described by the graphs of Fig. 14–9. In the p-type
material there is a large concentration N_p of positive carriers. These carriers are diffusing around and a
certain number of them each second approach the junction. This current of positive carriers which
approaches the junction is proportional to N_p. Most of them, however, are turned back by the high
potential hill at the junction and only the fraction $e^{-qV/\kappa T}$ gets through. There is also a current of
positive carriers approaching the junction from the other side. This current is also proportional to the
density of positive carriers in the n-type region, but the carrier density here is much smaller than the
density on the p-type side. When the positive carriers approach the junction from the n-type side, they
find a hill with a negative slope and immediately slide downhill to the p-type side of the junction. Let's
call this current I_0. Under equilibrium the currents from the two directions are equal. We expect then the
following relation:

$$I_0 \sim N_p(n\text{-side}) = N_p(p\text{-side}) e^{-qV/\kappa T}. \tag{14.12}$$

You will notice that this equation is really just the same as Eq. (14.10). We have just derived it in a different way.

Suppose, however, that we lower the voltage on the n-side of the junction by an amount ΔV — which we can do by applying an external potential difference to the junction. Now the difference in potential across the potential hill is no longer V but $V - \Delta V$. The current of positive carriers from the p-side to the n-side will now have this potential difference in its exponential factor. Calling this current I_1, we have

$$I_1 \sim N_p(p\text{-side}) e^{-q(V - \Delta V)/\kappa T}.$$

This current is larger than I_0 by just the factor $e^{q\Delta V/\kappa T}$. So we have the following relation between I_1 and I_0:

$$I_1 = I_0 e^{+q\Delta V/\kappa T}. \tag{14.13}$$

The current from the p-side increases exponentially with the externally applied voltage ΔV. The current of positive carriers from the n-side, however, remains constant so long as ΔV is not too large. When they approach the barrier, these carriers will still find a downhill potential and will all fall down to the p-side. (If ΔV is larger than the natural potential difference V, the situation would change, but we will not consider what happens at such high voltages.) The net current I of positive carriers which flows across the junction is then the difference between the currents from the two sides:

$$I_1 = I_0(e^{+q\Delta V/\kappa T} - 1). \tag{14.14}$$

The net current I of holes flows into the n-type region. There the holes diffuse into the body of the n-region, where they are eventually annihilated by the majority n-type carriers—the electrons. The electrons which are lost in this annihilation will be made up by a current of electrons from the external terminal of the n-type material.

When ΔV is zero, the net current in Eq. (14.14) is zero. For positive ΔV the current increases rapidly with the applied voltage. For negative ΔV the current reverses in sign, but the exponential term soon becomes negligible and the negative current never exceeds I_0 — which under our assumptions is rather small. This back current I_0 is limited by the small density of the minority p-type carriers on the n-side of the junction.

If you go through exactly the same analysis for the current of negative carriers which flows across the junction, first with no potential difference and then with a small externally applied potential difference ΔV, you get again an equation just like (14.14) for the net electron current. Since the total current is the sum of the currents contributed by the two carriers, Eq. (14.14) still applies for the total current provided we identify I_0 as the maximum current which can flow for a reversed voltage.
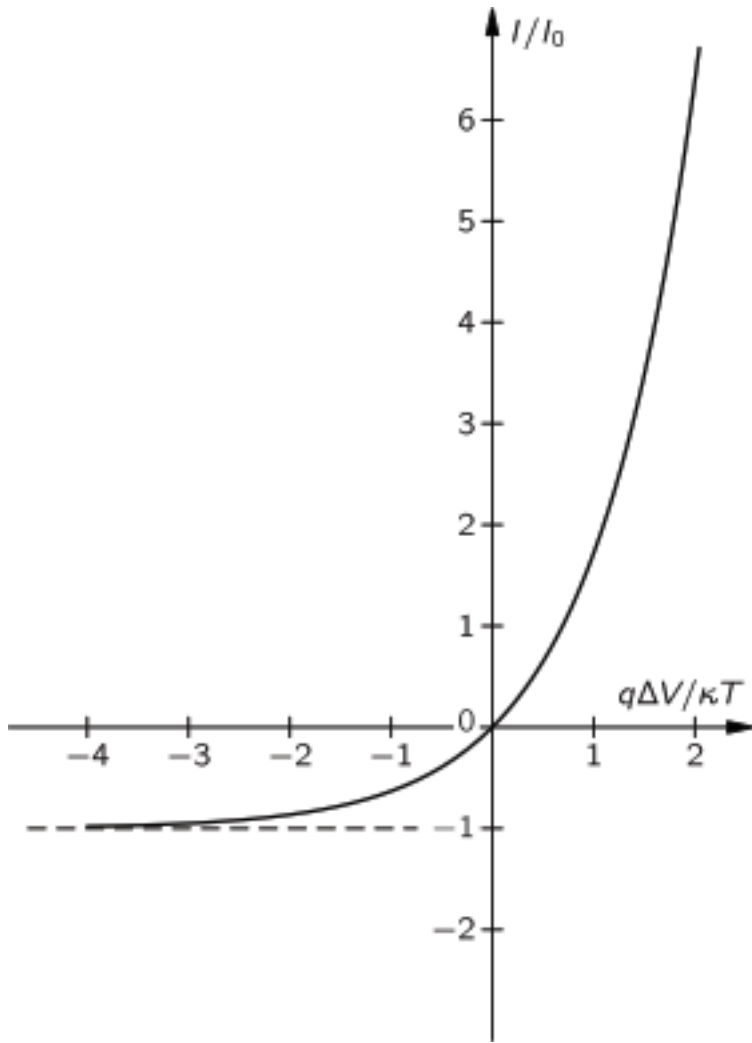
Fig. 14–10.The current through a junction as a function of the voltage across it.

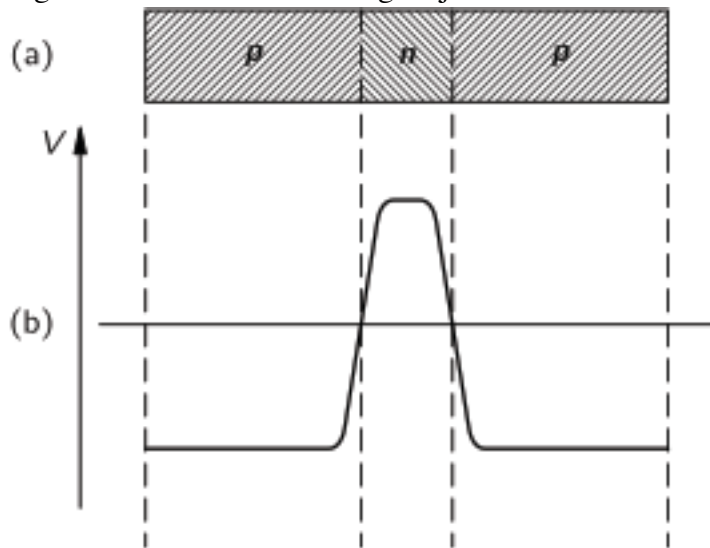
Fig. 14–11.The potential distribution in a transistor with no applied voltages.

The voltage-current characteristic of Eq. (14.14) is shown in Fig. 14–10. It shows the typical behavior of solid state diodes—such as those used in modern computers. We should remark that Eq. (14.14) is true only for small voltages. For voltages comparable to or larger than the natural internal voltage

difference V , other effects come into play and the current no longer obeys the simple equation.

14–6The transistor

Perhaps the most important application of semiconductors is in the transistor. The transistor consists of two semiconductor junctions very close together. Its operation is based in part on the same principles that we just described for the semiconductor diode—the rectifying junction. Suppose we make a little bar of germanium with three distinct regions, a p-type region, an n-type region, and another p-type region, as shown in Fig. 14–11(a). This combination is called a p-n-p transistor. Each of the two junctions in the transistor will behave much in the way we have described in the last section. In particular, there will be a potential gradient at each junction having a certain potential drop from the n-type region to each p-type region. If the two p-type regions have the same internal properties, the variation in potential as we go across the crystal will be as shown in the graph of Fig. 14–11(b).

Now let's imagine that we connect each of the three regions to external voltage sources as shown in part (a) of Fig. 14–12. We will refer all voltages to the terminal connected to the left-hand p-region so it will be, by definition, at zero potential. We will call this terminal the *emitter*. The n-type region is called the *base* and it is connected to a slightly negative potential. The right-hand p-type region is called the *collector*, and is connected to a somewhat larger negative potential. Under these circumstances the variation of potential across the crystal will be as shown in the graph of Fig. 14–12(b).
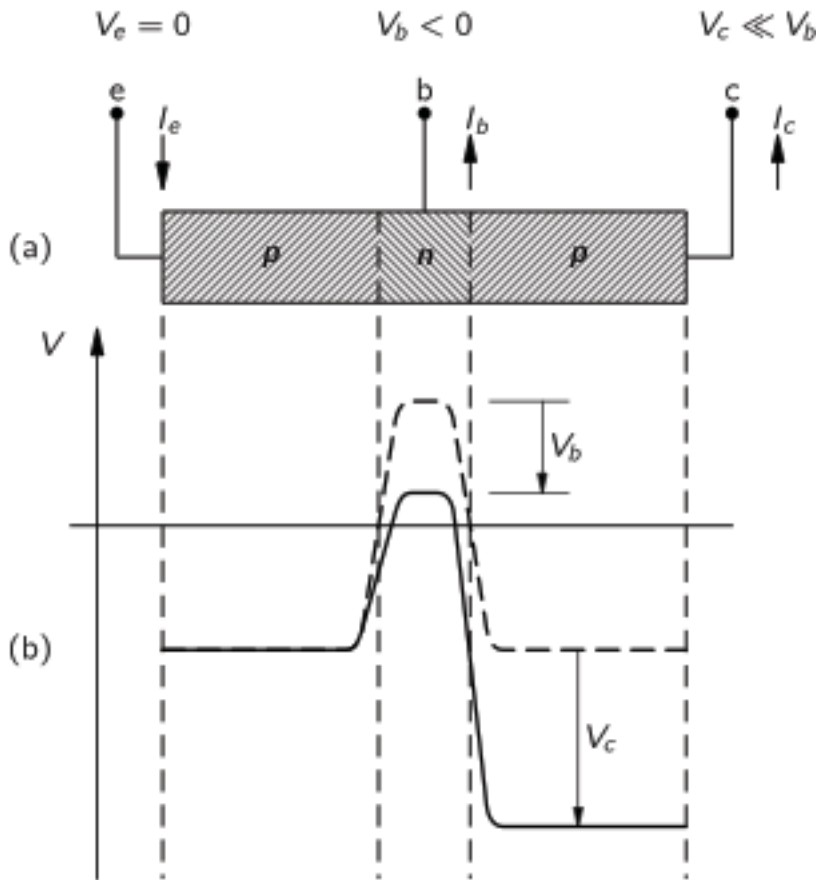
$V_e = 0$ $V_b < 0$ $V_c \ll V_b$

Fig. 14–12.The potential distribution in an operating transistor.

Let's first see what happens to the positive carriers, since it is primarily their behavior which controls the operation of the p-n-p transistor. Since the emitter is at a relatively more positive potential than the base, a current of positive carriers will flow from the emitter region into the base region. A relatively large current flows, since we have a junction operating with a "forward voltage"—corresponding to the right-hand half of the graph in Fig. 14–10. With these conditions, positive carriers or holes are being "emitted" from the p-type region into the n-type region. You might think that this current would flow out of the n-type region through the base terminal b. Now, however, comes the secret of the transistor. The n-type region is made very thin—typically 10^{-3} cm or less, much narrower than its transverse dimensions. This means that as the holes enter the n-type region they have a very good chance of diffusing across to the other junction before they are annihilated by the electrons in the n-type region. When they get to the right-hand boundary of the n-type region they find a steep downward potential hill and immediately fall into the right-hand p-type region. This side of the crystal is called the collector because it "collects" the holes after they have diffused across the n-type region. In a typical transistor, all but a fraction of a percent of the hole current which leaves the emitter and enters the base is collected in the collector region, and only the small remainder contributes to the net base current. The sum of the base and collector currents is, of course, equal to the emitter current.

Now imagine what happens if we vary slightly the potential V_b on the base terminal. Since we are on a relatively steep part of the curve of Fig. 14–10, a small variation of the potential V_b will cause a rather large change in the emitter current I_e. Since the collector voltage V_c is much more negative than the base voltage, these slight variations in potential will not effect appreciably the steep potential hill between the base and the collector. Most of the positive carriers emitted into the n-region will still be caught by the collector. Thus as we vary the potential of the base electrode, there will be a corresponding

variation in the collector current I_c. The essential point, however, is that the base current I_b always remains a small fraction of the collector current. The transistor is an amplifier; a small current I_b introduced into the base electrode gives a large current—100 or so times higher—at the collector electrode.

What about the electrons—the negative carriers that we have been neglecting so far? First, note that we do not expect any significant electron current to flow between the base and the collector. With a large negative voltage on the collector, the electrons in the base would have to climb a very high potential energy hill and the probability of doing that is very small. There is a very small current of electrons to the collector.

On the other hand, the electrons in the base *can* go into the emitter region. In fact, you might expect the electron current in this direction to be comparable to the hole current from the emitter into the base. Such an electron current isn't useful, and, on the contrary, is bad because it increases the total base current required for a given current of holes to the collector. The transistor is, therefore, designed to minimize the electron current to the emitter. The electron current is proportional to $N_n(\text{base})$, the density of negative carriers in the base material while the hole current from the emitter depends on $N_p(\text{emitter})$, the density of positive carriers in the emitter region. By using relatively little doping in the n-type material $N_n(\text{base})$ can be made much smaller than $N_p(\text{emitter})$. (The very thin base region also helps a great deal because the sweeping out of the holes in this region by the collector increases significantly the average hole current from the emitter into the base, while leaving the electron current unchanged.) The net result is that the electron current across the emitter-base junction can be made much less than the hole current, so that the electrons do not play any significant role in operation of the p-n-p transistor. The currents are dominated by motion of the holes, and the transistor performs as an amplifier as we have described above.

It is also possible to make a transistor by interchanging the p-type and n-type materials in Fig. 14–11. Then we have what is called an n-p-n transistor. In the n-p-n transistor the main currents are carried by the electrons which flow from the emitter into the base and from there to the collector. Obviously, all the arguments we have made for the p-n-p transistor also apply to the n-p-n transistor if the potentials of the electrodes are chosen with the opposite signs.