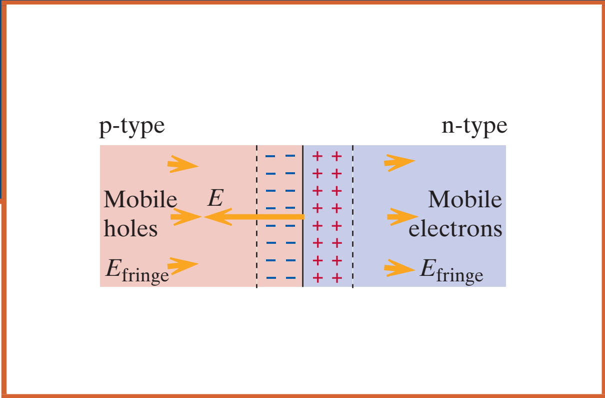
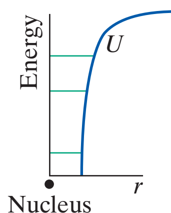
SUPPLEMENT

S2

Semiconductor Devices





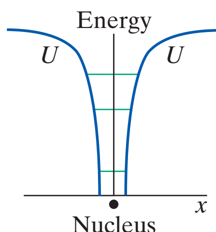


Figure S2.1 Quantized electronic energies in an isolated atom whose potential energy is U, shown as a function of r and as a function of x.

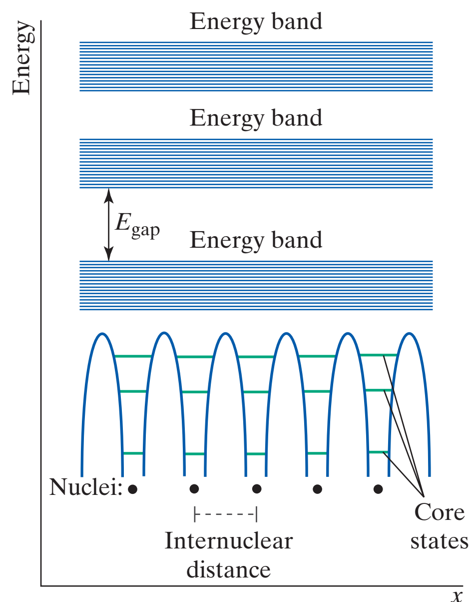


Figure S2.2 In a crystalline solid, the potential energy is periodic, and there are bands of closely spaced allowed energies.

OBJECTIVES

After studying this supplement you should be able to

- Apply Gauss's law to a semiconductor junction.

S2.1 SEMICONDUCTOR DEVICES

Computers and other modern electronic apparatus are built with semiconductor devices such as diodes and transistors. What is special about semiconductors such as silicon is that we can control the conductivity dynamically, by manipulating the number density of charge carriers. By applying Gauss's law we can understand some of the basic aspects of the design of semiconductor devices.

The simplest semiconductor device is a diode, which conducts current easily in one direction but conducts almost no current in the other direction. A diode can be made by joining two pieces of semiconducting material such as silicon, where one of the pieces contains a small amount of boron and the other contains a small amount of phosphorus. We will study the properties of such a semiconducting junction, which is a fundamental building block of modern electronic circuits. A particularly striking aspect of a diode is that there is an electric field "locked into" the interior of the material, which is something we've never dealt with before.

Understanding the behavior of a semiconducting junction is the foundation for understanding transistors, which are the most common element in digital circuits. Transistors can be constructed from two semiconducting junctions.

Energy Bands in Solids

The possible electronic energy states of an isolated atom are quantized. Figure S2.1 shows the electric potential energy U for an isolated atom, and the narrow energy levels. We have typically plotted U as a function of r, the separation of nucleus and electron, but in Figure S2.1 we also plot U as a function of x, with the nucleus at the origin, and electrons can be to the left or right of the origin.

If we assemble many such atoms to form a crystalline solid, the electric potential energy of the solid is periodic in x (and y and z) with a period corresponding to the distance between nuclei, as shown in Figure S2.2. In the case of such a periodic potential energy, quantum mechanics predicts—and many kinds of experiments confirm—that the energy-level scheme is quite different from that of a single atom. The higher energy states are thick bands of very closely spaced levels, often with large separations between these

bands of allowed energies. One of these separations is labeled E_{gap} in Figure S2.2. The large number of closely spaced energy levels corresponds to the large number of multielectron states in a solid.

Some electrons in the solid are bound to atomic cores, and electrons in these core states are not mobile. Electrons in an energy band are bound to the solid as a whole rather than to an individual atom, and such electrons can move relatively freely around in the solid if the energy band is not full, as we will see shortly.

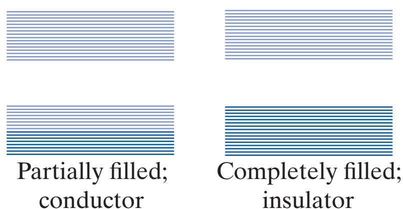


Figure S2.3 The highest occupied energy band is partially filled or completely filled.

It is an important property of electrons that no two electrons can be in exactly the same quantum state, a property that is referred to as the "Pauli exclusion principle." This means that even at very low temperatures, the electrons in an energy band will not all be crowded into the lowest energy state in that band. Rather, electrons will fill the band, with no more than two electrons in any one electronic state (the two electrons must differ in their spin directions in order to obey the Pauli exclusion principle).

If an energy band is full, other electrons must go into the next higher energy band. The highest energy band that contains electrons is either partially filled, which is the case for a conductor (a metal), or completely filled, which is the case for an insulator (Figure S2.3).

> **QUESTION** For the solid to be a conductor, it must be possible to add a small amount of energy to an electron that is at rest, so that it acquires some drift speed. Why will a solid be an insulator if its highest occupied energy band is completely filled?

The highest occupied energy band is separated from the next higher (unoccupied) energy band by the band gap energy E_{gap}. We can't add a small amount of energy to change an electron's energy within the energy band, because all the energy states are filled with two electrons each, and the Pauli exclusion principle forbids putting a third electron into the same energy state (Figure S2.4). In order to increase the energy of an electron in a completely filled energy band, we would need to provide a lot of energy, E_{gap} or more, to move the electron to an unoccupied level near the bottom of the higher, unoccupied energy band.

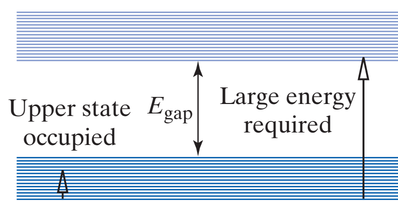


Figure S2.4 One cannot promote an electron from one energy state to another within the completely filled energy band of an insulator, and it takes a lot of energy to move to an empty higher energy band.

In some insulators with filled bands such as diamond, the band-gap energy is typically 5 electron volts (eV) or more. It is not possible to move an electron to a higher energy state without supplying a very large amount of energy. At ordinary temperatures, such a material is a very good insulator and a very poor conductor of electric current.

> **QUESTION** In contrast, why is the material a conductor if the energy band is only partially full?

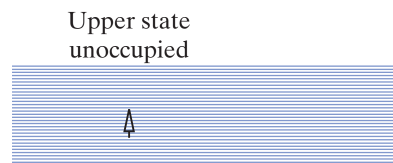


Figure S2.5 One can promote an electron from one energy state to another within the partially filled energy band of a conductor.

In a partially filled band we can add a small amount of energy to electrons near the top of the occupied levels, moving them to nearby unoccupied levels (Figure S2.5). As a result, the material is a conductor (a metal). The newly acquired extra energy can be lost in a collision, with the electron falling back down to a lower energy. This is an echo of the simple start–stop model of electron conduction in a metal.

Having moved an electron from a lower energy to a higher energy, there is a kind of "hole" in the electron sea of the conductor. Not only can the electron move (opposite to the applied electric field), but also the hole can move in the direction of the electric field, like a mobile positive particle. The effect is a bit like the motion of a bubble in water: the water moves down, but the bubble moves up. Unfortunately, a full description of the behavior of electrons and holes in a metal requires advanced solid state and quantum physics. The experimental facts are that most metals behave as though the mobile charges are negative electrons. A few metals, including aluminum and zinc, behave as though the mobile charges are positive holes.

In our simplest model of a metal whose mobile charges are electrons, there is a sea of mobile electrons with a background of positive atomic cores (ions). The drift of mobile electrons due to an applied electric field constitutes an electric current (Figure S2.6). In later discussions it will be important to distinguish between mobile charges and fixed (nonmobile) charges, so we will indicate mobile charges with circles and nonmobile charges without circles in diagrams such as Figure S2.6.

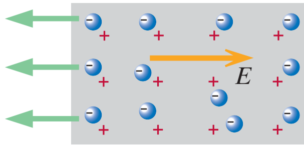


Figure S2.6 Electron conduction in a metal. There are mobile electrons and fixed positive atomic cores (ions). We will indicate mobile charges with circles; nonmobile charges have no circles.

Semiconductors

In a few materials with completely filled energy bands such as pure silicon or pure germanium, the band-gap energy is relatively small. Such materials are called semiconductors. The band-gap energy E_{gap} in silicon is 1.1 eV, and in germanium it is 0.7 eV. At room temperature these materials do conduct to some extent due to thermal effects. Although their intrinsic conductivity is low compared to a metal, we will see that it is possible to control the conductivity of a semiconductor dynamically, by manipulating the number density of mobile charges.

In a semiconductor, thermal energy can occasionally promote an electron into the next higher (unoccupied) energy band (Figure S2.7), and these electrons can pick up energy (and drift speed) from an applied electric field, because other, higher-energy states are nearby. The filled energy band is called the “valence” band, and the nearly unoccupied higher energy band is called the “conduction” band.

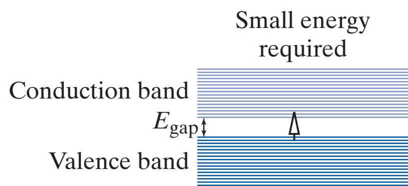


Figure S2.7 A semiconductor is a material with a small band gap.

Roughly, the number of electrons thermally excited into the conduction band is proportional to the Boltzmann factor e^{-E_{gap}/k_BT}, where k_B is the Boltzmann constant and k_BT at room temperature is about 1/40 eV:

$$k_BT \approx (1.4\times10^{-23}\,\text{J/K})(293\,\text{K})\left(\frac{1\,\text{eV}}{1.6\times10^{-19}\,\text{J}}\right) \approx 0.026\text{ eV} \approx \frac{1}{40}\text{ eV}$$

Even in germanium with its relatively small band-gap energy ($E_{\text{gap}} = 0.7$ eV) the Boltzmann factor is not very favorable: $e^{-(0.7\,\text{eV})/((1/40)\text{eV})} \approx e^{-28} \approx 1\times10^{-12}$. Therefore we might expect the conductivity of pure germanium at room temperature to be about 1×10^{-12} times smaller than the conductivity of an ordinary metal such as copper. A full analysis shows that the relevant factor is $e^{-E_{\text{gap}}/(2k_BT)}$, which is about 1×10^{-6}, which still yields a very low conductivity compared to a metal.

QUESTION In germanium or silicon, what will happen to the (low) electric conductivity if you raise the temperature?

When you raise the temperature, thermal energy promotes more electrons to the conduction band. Since the electron current $nA\bar{v} = nAuE$ is proportional to n, the number density of mobile electrons (number per unit volume), the electric conductivity of pure silicon or germanium increases with increasing temperature. In fact, the conductivity increases extremely rapidly with increasing temperature (Figure S2.8). This dramatic effect can be exploited in various ways, including using the current as an indicator of the temperature.

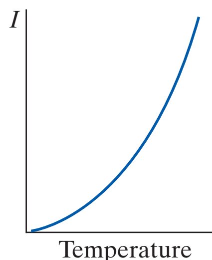


Figure S2.8 Current in a pure semiconductor grows extremely rapidly with temperature.

It is true that at higher temperatures the electron mobility decreases, associated with increased thermal motion of the atoms in the material. However, this modest decrease in electron mobility is swamped by the huge increase in the number density. The conductivity $\sigma = |q|nu$ depends on both the mobility u and the number density n, and σ actually increases dramatically.

The strongly temperature-dependent conductivity of pure silicon is called *intrinsic* conductivity. In a later section we will discuss another, different mechanism for conductivity in silicon that contains impurities (that is, some silicon atoms are replaced by other kinds of atoms).

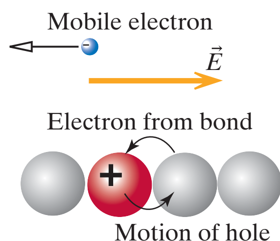


Figure S2.9 A mobile hole moving from atom to atom, to the right, in the direction of the applied electric field. Mobile electrons move to the left.

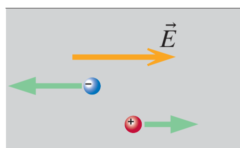


Figure S2.10 Intrinsic semiconductor conduction involves both holes and electrons.

Conduction by Holes When thermal energy happens to promote an electron to the conduction band in a semiconductor, the newly mobile electron drifts away, and an atom is now missing one electron. This atom has a net charge of $+e$, and we say that there is a "hole": there is a defective bond with a neighboring silicon atom. An electron in a neighboring atom associated with the bond could fill the hole. As a result, the location of the hole (the positive charge) could move from atom to atom, in the direction of the applied electric field (Figure S2.9). The effect is as though there are actual positive charges, mobile holes, that can move through the material.

If there are any mobile electrons, they drift through the material in a direction opposite to the electric field. Both the mobile electrons and the mobile holes contribute to the conventional current, in the direction of the electric field (Figure S2.10).

In a pure semiconductor, where thermal energy creates mobile electrons and mobile holes in pairs, the number density n_n of mobile (negative) electrons is equal to the number density n_p of mobile (positive) holes. In impure semiconductors these number densities need not be equal. Also, the mobilities of the electrons and holes are usually different.

In general, the total conventional current in a semiconductor is this:

$$I = en_n A u_n E + en_p A u_p E$$

Since $J = I/A = \sigma E$, the conductivity is $\sigma = en_n u_n + en_p u_p$.

Differences between Metals and Semiconductors

> **QUESTION** Summarize: What are the major differences between intrinsic conductivity in a semiconductor and conductivity in a metal?

In a metal, there are always lots of mobile charges, because the energy band is only partially filled, and changing the temperature hardly changes the number density of mobile charges.

In most metals there are essentially no mobile holes, and only the mobile electrons contribute to the current when an electric field is applied. Every atomic core is a positive ion, so there is no way for a neighboring atom to give up (another) electron and thereby create a mobile hole.

In contrast, in a pure semiconductor the (low) densities of mobile electrons and holes depend strongly on the temperature, and both electrons and holes contribute to the conductivity. The fraction of atomic cores that are positive ions is very small.

Other Pair-Creation Mechanisms For completeness, we note that thermal excitation is not the only mechanism for creating mobile electron–hole pairs in a semiconductor. Absorption of a photon can create a pair, and this is the basis for photovoltaic devices (solar cells).

In another application, the passage of a high-energy charged particle through a semiconductor creates a number of electron-hole pairs that is proportional to the energy loss of the particle, so a semiconductor can be used as a detector of charged particles. Collecting the mobile charges gives a measure of the energy loss of the particle. If the semiconductor is large enough that the particle comes to a stop inside the material, the measured energy loss represents the original kinetic energy of the particle.

Recombination Sometimes one of the mobile electrons in a semiconductor happens to come near a hole (that is, an atom that is missing an electron and therefore has a charge of $+e$). The mobile electron is of course attracted to the atom and can fill the hole. This reduces both the number of mobile electrons and the number of mobile holes. We call this process recombination.

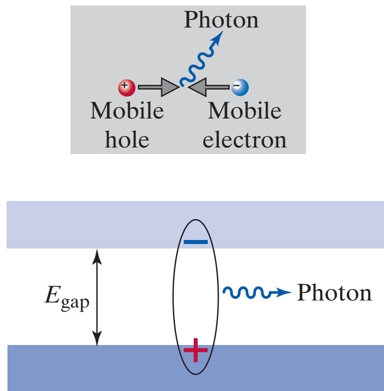


Figure S2.11 In an LED, the recombination of electrons and holes produces photons.

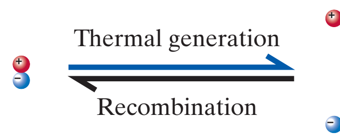


Figure S2.12 Recombination of electrons and holes competes with the thermal generation of electron–hole pairs. In the steady state the two rates are equal.

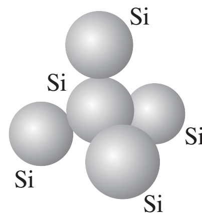


Figure S2.13 Silicon atoms bond to four neighboring silicon atoms in a tetrahedral arrangement.

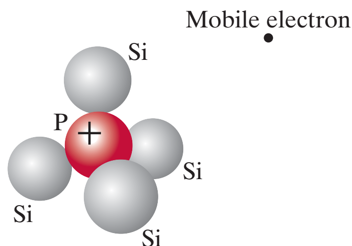


Figure S2.14 A silicon atom is replaced by a phosphorus atom, which contributes a mobile electron.

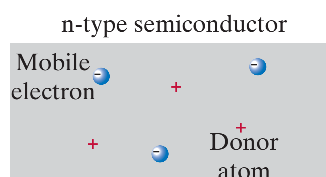


Figure S2.15 An n-type semiconductor, doped with a small number of fixed donor atoms that donate mobile electrons.

This process is similar to the recombination of ions and electrons in a spark or plasma, where recombination creates a bound-state atom accompanied by the emission of light, whose energy corresponds to the binding energy of the atom. Because electron–hole recombination takes place inside a solid, the electron–hole binding energy need not be released as light but may be absorbed by neighboring atoms. However, it is possible to make a semiconductor device in which recombination results in the emission of light, with energy close to the band-gap energy (Figure S2.11). Such a device is called a light-emitting diode, or LED.

Recombination, which destroys electron–hole pairs, competes with pair creation by thermal excitation (Figure S2.12). In the steady state, the recombination rate becomes equal to the pair creation rate.

> **QUESTION** What happens to the number densities if the temperature is increased?

If the temperature is increased, the pair creation rate goes up, and for a while there is a growth in the number densities of mobile electrons and mobile holes. However, with higher number densities there is a higher probability of an electron encountering a hole and recombining, so eventually a new steady state is established at the higher temperature, with higher but constant densities of mobile electrons and mobile holes.

Summary of Types of Material

- Metals have a fixed high density of mobile electrons, and high conductivity. There are no mobile holes.
- Insulators have hardly any mobile electrons or mobile holes and have very low conductivity.
- In a pure semiconductor the number density of mobile electrons and mobile holes depends strongly on the temperature. The intrinsic conductivity of pure semiconductors grows extremely rapidly with increasing temperature, but it is only about 1×10^{-6} that of a metal at room temperature. Both mobile electrons and mobile holes contribute to the conductivity.

Doped Semiconductors

In a metal the number density of mobile charges is fixed. We can't vary this density. In a semiconductor, however, there is a possibility of changing the number densities of mobile electrons or holes, which opens up new possibilities for designing devices with special capabilities. As we have seen, one way to vary the mobile charge densities is to change the temperature in a pure semiconductor. Another way is to deliberately introduce impurities that can contribute mobile charges. This is called "doping."

An n-Type Semiconductor In pure silicon, each silicon atom has four outer electrons that form chemical bonds with outer electrons in four neighboring silicon atoms in a tetrahedral arrangement (Figure S2.13). Suppose that we replace one silicon atom with a phosphorus atom (Figure S2.14). A phosphorus atom is about the same size as a silicon atom, but it has five outer electrons instead of four (see the periodic table on the inside front cover of this textbook). If we replace a silicon atom with a phosphorus atom, four of the outer electrons of the phosphorus atom participate in bonds with the four neighboring silicon atoms. The fifth electron is only very weakly bound to the phosphorus atom and is very easily promoted into the conduction band by thermal excitation, leaving behind a positively charged phosphorus ion (Figure S2.15). The phosphorus atom is called a "donor" impurity (because it "donates" a mobile electron).

QUESTION Does the presence of a phosphorus atom generate a mobile hole?

The positive charge of the phosphorus ion does not constitute a mobile hole. The four outer electrons of each of the neighboring silicon atoms are participating fully in chemical bonds and are not available to fill the hole in the phosphorus atom. This is unlike the case in pure silicon, where a hole in a silicon atom means there is a defective bond with a neighbor, and an electron from a neighboring atom can fill the hole to repair the bond, with the effect of moving the hole.

Doping pure silicon by replacing even a very small fraction of the atoms with phosphorus atoms makes a material in which there are many more mobile electrons than in pure silicon. Such a material is called an “n-type semiconductor” (“n” stands for “negative”). With a typical level of doping of one phosphorus atom in tens of thousands of silicon atoms, at room temperature the mobile electrons donated by the phosphorus atoms are far more numerous than the electrons resulting from thermal creation of electron–hole pairs.

Majority Carriers in n-Type Semiconductors Are Electrons Doping with phosphorus greatly reduces the number of mobile holes. Flooding the silicon with mobile electrons means that mobile holes much more frequently encounter mobile electrons and recombine. We say that in an n-type semiconductor the “majority” mobile charges are electrons, and there are few “minority” mobile charges (holes).

It is important to keep in mind that the n-type semiconductor is electrically neutral overall. There are just as many positively charged phosphorus ions as there are mobile electrons. This is similar to an ordinary metal, where the positive charge of the atomic cores balances the negative charge of the mobile electrons (typically one for each atom).

A Dilute Metal You can think of an n-type semiconductor as a dilute metal. In both materials there are fixed positive ions and mobile electrons. The mobile electrons have little net force acting on them, and they move around quite freely in the material, which is neutral when averaged over volumes containing many positive atomic cores. The main difference is that in an ordinary metal there are one or more mobile electrons per atom, whereas in an n-type semiconductor only a small fraction of atoms give up an electron. The donor atoms are rather far apart, compared to the situation in a true metal where every atom can be an electron donor.

A p-Type Semiconductor It may seem that an n-type semiconductor isn’t very interesting or useful, since it is basically just like a metal but with a much smaller density of mobile electrons. After we discuss a somewhat more unusual material, a p-type semiconductor in which the majority mobile charges are holes rather than electrons, we will be able to discuss a “p-n junction,” consisting of p-type and n-type semiconductors right next to each other. P-n junctions have interesting properties and are a basic building block of transistors and integrated circuits.

The analysis of a p-type semiconductor is quite parallel to the analysis of an n-type semiconductor, and the next few paragraphs should sound familiar. A boron atom is about the same size as a silicon atom, but it has only three outer electrons instead of four (see the periodic table on the inside front cover of the textbook). If we replace a silicon atom with a boron atom, the boron atom grabs an additional electron in order to participate in bonds with the four neighboring silicon atoms (Figure S2.16). This leaves a hole in a neighboring silicon atom, and the hole is mobile because an electron from another silicon atom can fill this hole.

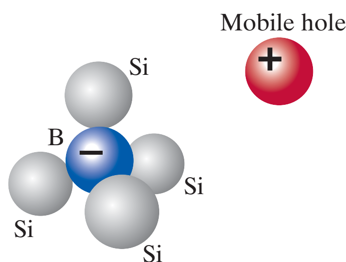


Figure S2.16 A silicon atom is replaced by a boron atom, which grabs an extra electron and creates a mobile hole.

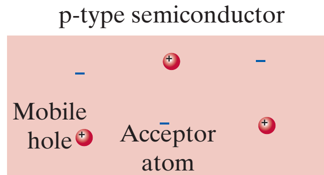


Figure S2.17 A p-type semiconductor, doped with a small number of fixed acceptor atoms that accept electrons, thereby creating mobile holes.

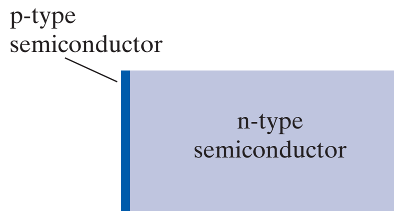


Figure S2.18 An n-type semiconductor with a p-type region introduced by ion implantation.

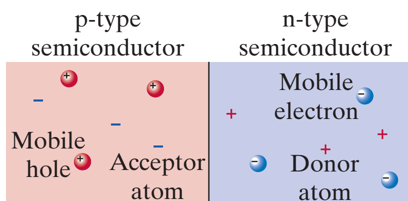


Figure S2.19 Close-up of a p-n junction. The majority mobile charges are holes in the p-type region and electrons in the n-type region.

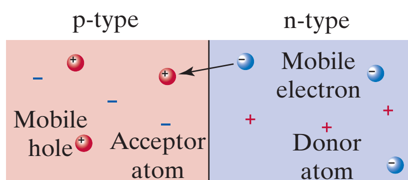


Figure S2.20 If an electron drifts into the p-type region, it almost immediately recombines with a hole.

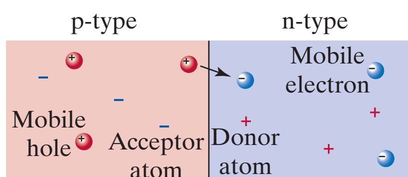


Figure S2.21 If a hole drifts into the n-type region, it almost immediately recombines with an electron.

The effect is to create a mobile hole and to leave behind a negatively charged boron ion (Figure S2.17). The boron atom is called an "acceptor" impurity (because it "accepts" an electron, thereby contributing a mobile hole). The negative charge of the boron ion obviously does not constitute a mobile electron.

Doping pure silicon by replacing even a very small fraction of the atoms with boron atoms makes a material in which there are many more mobile holes than in pure silicon. Such a material is called a "p-type semiconductor" ("p" stands for "positive"). With a typical level of doping of one boron atom in tens of thousands of silicon atoms, at room temperature the mobile holes contributed by the boron atoms are far more numerous than the holes resulting from thermal creation of electron–hole pairs.

Majority Carriers in p-Type Semiconductors Are Holes Doping with boron greatly reduces the number of mobile electrons. Flooding the silicon with mobile holes means that mobile electrons much more frequently encounter mobile holes and recombine. We say that in a p-type semiconductor the "majority" mobile charges are holes, and there are few "minority" mobile charges (electrons).

It is important to keep in mind that the p-type semiconductor is electrically neutral overall. There are just as many negatively charged boron ions as there are mobile holes.

How Doping Is Done To prepare n-type or p-type silicon, a crystal can be grown from molten silicon that contains precise amounts of donor or acceptor atoms. Another way to dope silicon is to accelerate donor or acceptor ions by electric fields and shoot them into the silicon at high speed. This technique is called "ion implantation" and is used to produce shallow regions of doped material.

Figure S2.18 shows n-type material into which a p-type region has been introduced by ion implantation. This is called a "p-n junction," which we will discuss next.

A p-n Junction

A rectifying diode is an extremely useful device that allows current to run through it in one direction but not the other. Figure S2.19 shows a close-up of a p-n junction, which acts like a diode. We will show that conventional current can run easily from the p-type section of the diode toward the n-type section, but very little current can flow in the opposite direction. The following diagrams represent the essential elements of a p-n junction, but the actual geometry may look rather different in real devices.

Keep in mind that in the p-type material there are lots of mobile holes and few mobile electrons. In the n-type material there are lots of mobile electrons and few mobile holes.

Consider what happens when a mobile electron happens to drift across the junction into the p-type material (Figure S2.20). It has now entered a region that is full of mobile holes, so it may recombine soon after crossing the junction. This removes a hole from circulation in the p-type material, leaving some nearby negatively charged acceptor atom unpaired with a mobile hole. It also leaves an unpaired donor atom in the n-type region.

Similarly, a mobile hole that happens to drift into the n-type material finds itself in a region that is full of mobile electrons, so it may recombine soon after crossing the junction (Figure S2.21). This removes an electron from circulation in the n-type material, leaving some nearby positively charged donor atom unpaired with a mobile electron. It also leaves an unpaired acceptor atom in the p-type region.

The result is that a thin region near the junction becomes depleted of mobile electrons and holes. This is called the "depletion layer" (Figure S2.22).

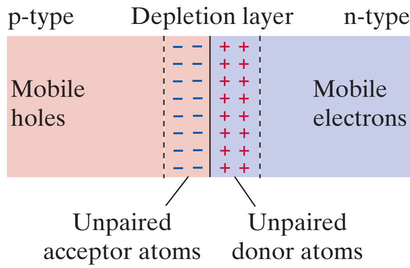


Figure S2.22 In the depletion layer there are unpaired acceptor and donor atoms, and few mobile holes or electrons.

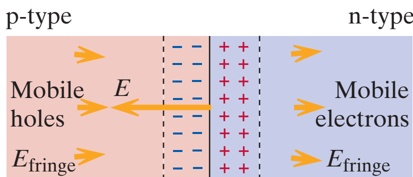


Figure S2.23 Pattern of electric field produced by the depletion layer.

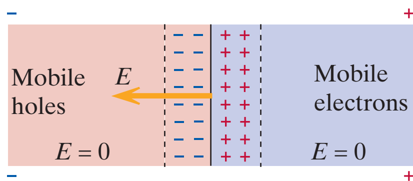


Figure S2.24 The diode polarizes, making $E = 0$ in the p-region and n-region.

The p-type part of the depletion layer has a negative charge due to the presence of the unpaired acceptor atoms (and lack of compensating mobile holes). The n-type part of the depletion layer has a positive charge due to the presence of the unpaired donor atoms (and lack of compensating mobile electrons). As we will see later, the depletion layer is very thin, typically only a few hundred atoms in width.

This gives rise to a strong electric field in the depletion layer as shown in Figure S2.23. Outside the depletion layer, the charged regions of the depletion layer produce a tiny "fringe field" rather like that outside the plates of a capacitor. This fringe field is very small because, as we will see later, the width of the depletion layer is only a few hundred atoms, and the fringe field of a capacitor is proportional to the spacing between plates.

The tiny fringe field drives holes to the right in the p-type material and drives electrons to the left in the n-type material. This polarizes the diode, leading to a tiny amount of charge on distant outer surfaces as shown in Figure S2.24. These charges are bare, positive donor atoms on the surface of the n-type material and bare, negative acceptor atoms on the surface of the p-type material.

> **QUESTION** In equilibrium, why must the net electric field in the material outside the depletion layer be zero?

Outside the depletion layer there are abundant mobile electrons (in the n-type material) or mobile holes (in the p-type material). As long as there is a nonzero electric field, these mobile charges will move, and they move in a direction so as to polarize the material in such a way as to lead to a zero net field. This is another example of the fact that inside a conductor in equilibrium the net electric field must become zero through polarization.

> **QUESTION** How can there be a nonzero electric field in the depletion layer?

The situation in the depletion layer is different. There are almost no mobile charges in this region, and in equilibrium there can be a strong electric field, which we will calculate in the next section.

The tiny amount of charge on the outer surfaces plays no significant role in the behavior of the p-n junction, so we will omit showing these charges in later diagrams.

The Electric Field in the Depletion Layer

We can use Gauss's law to calculate the nonzero electric field in the depletion layer of the p-n junction. We draw an (imaginary, mathematical) box with one end in the p-type material and the other in the depletion layer (Figure S2.25).

> **QUESTION** On which portions of the Gaussian box is there any nonzero electric flux? How much?

The electric field on the left end of the box is zero, because it is in the region where we know the electric field is zero in equilibrium. The electric field on the top, bottom, and sides is either zero (because it is in a zero-field region) or is horizontal (due to the symmetry of the situation), so the flux is zero on all of these sides. Only on the right end of the Gaussian box is there any flux, and this flux is $-EA$, where E is the (unknown) electric field, A is the area of the right end of the box, and the flux is negative because the electric field points into the box.

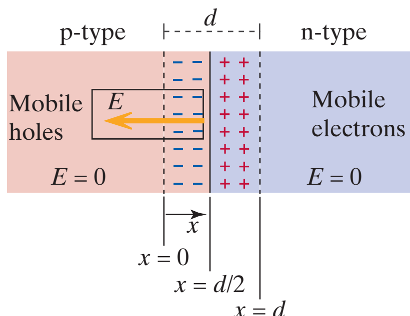


Figure S2.25 A Gaussian box with its right end (with area A) in the depletion layer. Measure x from the left edge of the depletion layer.

Choose a coordinate system with the origin at the left end of the depletion layer, whose width is d (Figure S2.25). We'll consider the simple case that the doping density is the same in the p-type and n-type materials, so that the doping charge density is $-\rho$ (in C/m^3) in the left portion and $+\rho$ in the right portion.

> **QUESTION** Use Gauss's law to calculate the electric field as a function of x, the location of the right end of the Gaussian box (measured from the left end of the depletion layer). It may be easiest to consider the left half of the depletion layer separately from the right half.

When $x < d/2$ (x is in the left portion of the depletion layer), we have the following, because the volume containing charge is Ax:

$$-EA = \frac{Q_{\text{inside}}}{\varepsilon_0} = \frac{-\rho(Ax)}{\varepsilon_0}$$
$$E = \frac{\rho x}{\varepsilon_0}$$

When $x > d/2$ (x is in the right portion of the depletion layer), we have the following:

$$-EA = \frac{Q_{\text{inside}}}{\varepsilon_0} = \frac{-\rho(Ad/2)+\rho[A(x-d/2)]}{\varepsilon_0}$$
$$E = \frac{\rho(d-x)}{\varepsilon_0}$$

Figure S2.26 is a graph of E vs. x through the depletion layer.

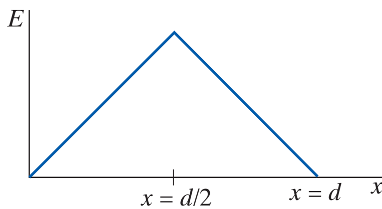


Figure S2.26 The magnitude of the electric field as a function of x, measured from the left edge of the depletion layer.

Dielectric Constant There is an additional factor that we have omitted from the analysis. Inside this dense material, the electric field contributed by a single bare charge (a donor or acceptor atom) is reduced by a factor K, the dielectric constant (see the end of Chapter 16 for a related discussion). This effect is related to the polarization of the material due to these embedded bare charges. As a result, the effective electric field inside the material is $1/K$ as large as the field we have calculated. In silicon this is a sizable effect; K is approximately 12.

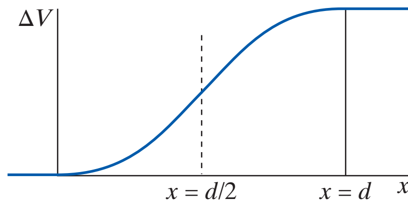


Figure S2.27 A graph of the increasing potential starting from the left edge of the depletion layer.

The Potential Difference Across the Depletion Layer

Since the electric potential difference is given by $\Delta V = -\int_i^f \vec{E}\bullet d\vec{l}$, the potential rises in going left to right through the depletion layer. Figure S2.27 shows the potential as a function of x obtained by integrating the triangle graph of electric field E in Figure S2.26. The result is two parabolas. The total potential difference across the depletion layer is the area of the triangle in Figure S2.26. Taking account of the dielectric constant K, we have this:

$$\Delta V = \frac{d}{2}E_{\text{max}} = \frac{\rho d^2}{4K\varepsilon_0}$$

Could we use this locked-in potential difference to drive a perpetual current in a circuit? Not surprisingly, the answer is no. For example, suppose that you extend the p-type material into a long wire that comes around and attaches to the end of the n-type material to form a complete circuit. Now in addition to the p-n junction there is an n-p junction, with a locked-in potential difference in the opposite direction. The round-trip path integral of the electric field is zero, and there is no emf to drive current.

If you connect a metal wire across the p-n junction, a similar situation arises. Although we have not discussed the issue, perhaps you can see that at the junction between a metal and a semiconductor there can be a depletion layer with an associated potential difference.

The Band-Gap Energy If ΔV is the potential difference across the depletion layer, we can show that $e\Delta V$ is equal to the band-gap energy E_{gap}. To create a

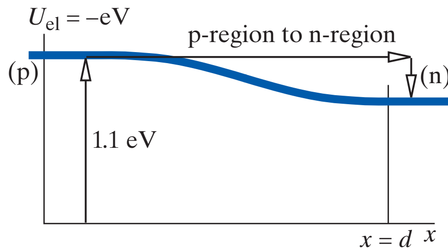


Figure S2.28 A freed-up electron in the p-region falls into the n-region.

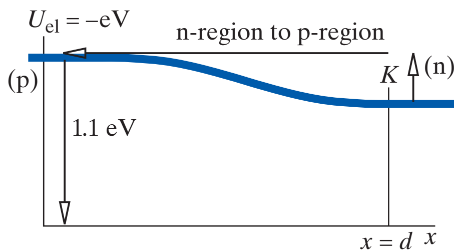


Figure S2.29 An already free electron in the n-region gains enough kinetic energy K to move into the p-region.

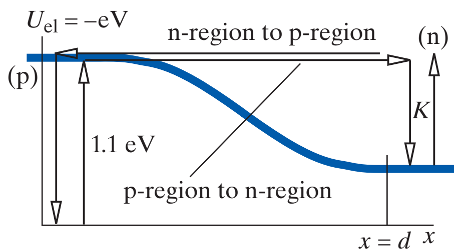


Figure S2.30 Transfer of electrons from the n-region to the p-region increases the potential energy barrier.

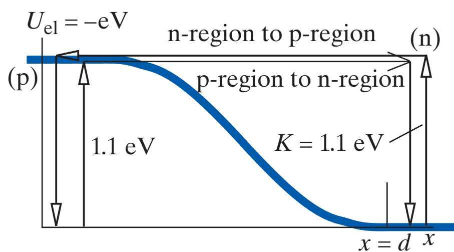


Figure S2.31 A dynamic equilibrium is achieved when the kinetic energy K required to cross the barrier is equal to the band-gap energy.

mobile electron in a semiconductor, one must supply the band-gap energy to promote the electron into an energy level corresponding to a mobile electron.

In Figure S2.28 we plot the electric potential energy $U_{el} = -eV$ of an electron in a silicon diode. The sign of the curve is opposite to that in Figure S2.27, because we are plotting the potential energy U_{el} for a negative electron instead of the potential V. Figure S2.28 refers to a time shortly after the depletion layer begins to form. It takes 1.1 eV of thermal excitation to create a mobile electron in the p-region of the silicon. Once mobile, this electron can "fall" through the potential energy difference into the n-region.

In contrast, at this time an already mobile electron in the n-region in Figure S2.29 needs to obtain very little kinetic energy K from thermal excitation to be able to cross into the p-region, where it can drop 1.1 eV to bind with a silicon atom that is missing an electron (a hole). The smaller energy requirement means that initially it is much more likely for electrons to move from the n-region to the p-region than from the p-region to the n-region.

The net flow of electrons from the n-region to the p-region increases the potential energy difference, as shown in Figure S2.30. The increased energy barrier causes a decrease in the rate of electrons going from the n-region to the p-region, making the net flow smaller. The depletion layer will grow until the potential energy difference is equal to 1.1 eV, at which point there is a dynamic equilibrium and no more growth of the depletion layer.

In Figure S2.31 we plot the electric potential energy in the dynamic equilibrium situation. It takes 1.1 eV of thermal excitation to create a mobile electron in the p-region. Once free, this electron can "fall" through the potential energy difference into the n-region. Similarly, an already-mobile electron in the n-region must obtain kinetic energy K of 1.1 eV of thermal excitation to be able to cross into the p-region, where it can drop 1.1 eV to bind with a silicon atom that is missing an electron (a hole). There is no net advantage to moving in one direction or the other. In dynamic equilibrium the two processes have equal rates.

In terms of the Boltzmann factor, the rate for creating a mobile electron is proportional to $e^{-E_{gap}/k_B T}$, where ΔE is 1.1 eV for silicon. The rate for electrons in the n-region to obtain kinetic energy K from thermal excitation is proportional to $e^{-K/k_B T}$. When $K = \Delta E$ the two rates are equal.

Numerical Values

We can use these results to show that the depletion layer has a width of just a few hundred atoms. We can also calculate the magnitude of the electric field in the depletion layer.

Using the measured value for the band-gap energy, we can write the following for silicon:

$$E_{gap} = e\Delta V = \frac{e\rho d^2}{4K\varepsilon_0} = (1.1\,\text{eV})\left(\frac{1.6\times10^{-19}\,\text{J}}{\text{eV}}\right)$$

Since $e = 1.6\times10^{-19}$ C we have

$$\Delta V = \frac{\rho d^2}{4K\varepsilon_0} = 1.1\,\text{J/C} = 1.1\,\text{V}$$

We can use this result to determine the width d of the depletion layer if we calculate the charge density ρ. A typical doping level in silicon is 1×10^{18} atoms per cm^3. The density of silicon is 2.4 g/cm^3, and there are 28 g per mole, so

$$\left(\frac{6\times10^{23}\text{ atoms}}{28\text{ g}}\right)\left(\frac{2.4\text{ g}}{\text{cm}^3}\right) = 5\times10^{22}\text{ silicon atoms/cm}^3$$

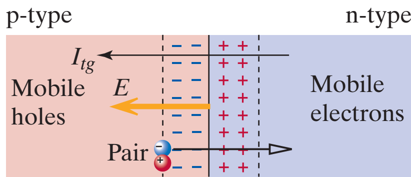


Figure S2.32 Electrons thermally generated on the left are swept through the depletion layer to the right, contributing to the thermal-generation conventional current I_{tg} to the left. $K_i \approx 0$ is sufficient.

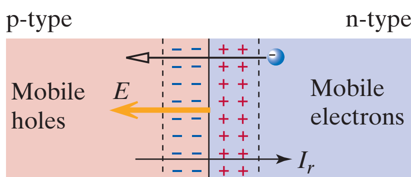


Figure S2.33 Occasionally an electron has enough energy to move from right to left against the electric force, contributing to the recombination conventional current I_r to the right. We need $K_i \approx E_{\text{gap}}$.

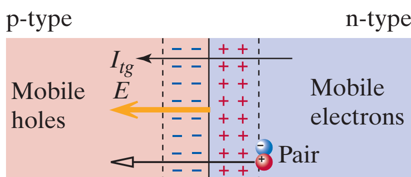


Figure S2.34 Holes thermally generated on the right are swept through the depletion layer to the left, contributing to the thermal-generation conventional current I_{tg} to the left. $K_i \approx 0$ is sufficient.

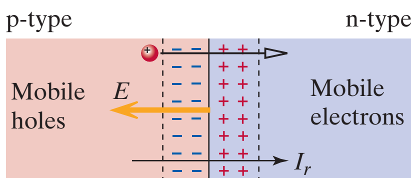


Figure S2.35 Occasionally a hole has enough energy to move from left to right against the electric force, contributing to the recombination conventional current I_r to the right. We need $K_i \approx E_{\text{gap}}$.

In other words, a typical doping is about one impurity atom per 50,000 silicon atoms. The charge density of the donor or acceptor atoms is

$$\rho = (1.6 \times 10^{-19}\,\text{C})\left(\frac{1 \times 10^{18}\text{ atoms}}{1 \times 10^{-6}\,\text{m}^3}\right) = 1.6 \times 10^5\,\text{C/m}^3$$

$$d = \sqrt{\frac{4K\varepsilon_0 \Delta V}{\rho}} = \sqrt{\frac{(4)(12)(9 \times 10^{-12}\,\text{C}^2/\text{N}\cdot\text{m}^2)(1.1\,\text{J/C})}{(1.6 \times 10^5\,\text{C/m}^3)}} = 550 \times 10^{-10}\,\text{m}$$

The depletion layer is very narrow, roughly 500 atoms wide. Now that we have a value for d, we can find the maximum electric field:

$$\Delta V = \frac{d}{2} E_{\text{max}}$$

so

$$E_{\text{max}} = 2\frac{\Delta V}{d} = \frac{2(1.1\text{V})}{550 \times 10^{-10}\,\text{m}} = 4 \times 10^7\,\text{V/m}$$

This is a very large electric field. For example, it is over 10 times the field that causes air to ionize (about 3×10^6 V/m). On the other hand, it is very small compared to the field needed to ionize an isolated atom, which is about 1×10^{11} V/m.

Diode Action

Now we can explain how a sizable conventional current can run from the p-region toward the n-region, but only a very small current can run through the junction in the other direction. Such a device is called a diode.

In equilibrium, there is no current running from the p-region to the n-region, or vice versa. At the atomic level this is a dynamic equilibrium, not a static one. Occasionally thermal energy creates an electron–hole pair in the p-region (Figure S2.32). If the electron survives long enough to wander into the depletion layer (or is created in the depletion layer), it is quickly swept across the junction by the electric field, and it joins the many mobile electrons in the n-region. This contributes to a "thermal-generation" conventional current I_{tg} to the left.

Occasionally one of the abundant mobile electrons in the n-region happens through random thermal collisions to have enough kinetic energy to move through the depletion layer, against the electric force $(-e)\vec{E}$ (Figure S2.33). When the electron reaches the p-region it recombines with one of the abundant mobile holes found there. This contributes to a "recombination" conventional current I_r to the right.

QUESTION Can you think through these two processes with respect to holes?

Similarly, a hole that is thermally generated in the n-region may survive long enough to wander into the depletion layer (or it may be created in the depletion layer), where it is quickly swept across the junction by the electric field, and it joins the many mobile holes in the p-region (Figure S2.34). This too contributes to the thermal-generation conventional current I_{tg} to the left.

Occasionally one of the abundant mobile holes in the p-region happens through random thermal collisions to have enough kinetic energy to move through the depletion layer, against the electric force $(+e)\vec{E}$ (Figure S2.35). When the hole reaches the n-region it recombines with one of the abundant mobile electrons found there. This contributes to the recombination conventional current I_r to the right.

In (dynamic) equilibrium the two processes are balanced: $I_{tg} = I_r$, and the net conventional current to the right $I = I_r - I_{tg}$ is zero. We can create an

imbalance in the two types of current and generate a nonzero net current by altering the potential difference across the diode with a battery.

"Reverse" Voltage and Current Suppose that we connect a battery to the diode as shown in Figure S2.36. This arrangement applies what is called a "reverse" voltage to the diode. We will see that this leads to a nonzero but small current through the diode.

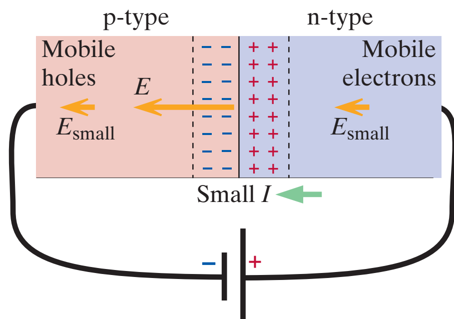


Figure S2.36 Applying a "reverse" voltage to the diode produces a small conventional current to the left.

With the + end of the battery attached to the n-region, the potential difference ΔV across the diode is increased. The depletion layer becomes somewhat thicker. One way to visualize what happens is to see that the battery will pull some electrons out of the n-region, leaving a thicker positive portion of the depletion layer. Similarly, the battery pulls holes out of the p-region, leaving a thicker negative region of the depletion layer. A thicker depletion layer produces a larger average electric field in the layer, which creates a larger potential difference across the layer.

> **QUESTION** Even if a net current runs through the diode, most of this potential difference will be across the depletion layer. Do you see why?

Remember that the bulk p- and n-regions have lots of mobile charges (holes in the p-region, electrons in the n-region), whereas there are few mobile charges in the depletion layer. In the steady state the electric field in the bulk material will be very small compared to the electric field in the depletion layer, since $I = neAuE$ must be the same throughout the diode. Large number density n means small electric field E. Therefore most of the potential difference appears across the depletion layer.

Now that the potential difference is larger across the depletion layer, it is significantly less likely that an electron in the n-region or a hole in the p-region will happen to have enough kinetic energy to move through the depletion layer against the potential difference.

> **QUESTION** As a result of the increased potential difference, what will happen to I_{tg}? To I_r?

The thermal-generation current I_{tg} won't change. It is made up of electrons and holes that are swept quickly through the depletion layer by the electric field anyway, and they will continue to be swept through quickly. A small initial kinetic energy $K_i \approx 0$ is sufficient.

The recombination current I_r, however, is made up of electrons and holes that had to travel against the electric force, and now it is even harder to go against the larger forces corresponding to the larger potential difference. A large initial kinetic energy $K_i \approx E_{gap}$ is needed. Therefore the recombination current will decrease rapidly with small increases in potential difference across the depletion layer and may be very small compared to the thermal-generation current. (In terms of the Boltzmann factor, the recombination current decreases by an amount $e^{-e|\Delta V_{batt}|/k_B T}$, where ΔV_{batt} is the additional potential difference due to the battery.)

> **QUESTION** What is the direction of the net current, which is approximately I_{tg}, since I_r is now very small?

The thermal-generation conventional current flows in the direction of the electric field that is in the depletion layer, so it flows to the left, from the n-type region to the p-type region, in the expected direction for the way the battery is attached.

It is important to understand that this nonzero current is small in an absolute sense because the thermal-generation current is essentially what you would obtain in pure silicon. The mobile electrons and holes contributed by the donor and acceptor atoms play little role. Incidentally, the magnitude of

this small current is sensitive to light hitting the junction, since the absorption of light is another way to generate electron–hole pairs. This is the basis for photocells.

After increasing the potential difference enough to make the recombination current negligible, further increases in the potential difference lead to essentially no change in the small current.

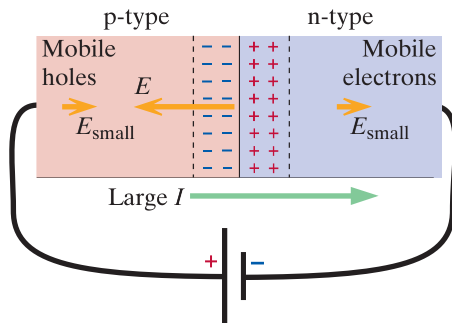


Figure S2.37 Applying a "forward" voltage to the diode produces a large conventional current to the right.

"Forward" Voltage and Current Next we connect a battery to the diode as shown in Figure S2.37. This arrangement applies what is called a "forward" voltage to the diode. We will see that this battery connection leads to a large current through the diode.

With the + end of the battery attached to the p-region, the potential difference ΔV across the diode is decreased. Most of this decreased potential difference appears across the depletion layer, since the current $neAuE$ must be the same throughout the diode. One way to visualize what happens is to see that the battery will push some extra electrons into the n-region, leading to a thinner positive portion of the depletion layer. Similarly, the battery pushes holes into the p-region, leading to a thinner negative region of the depletion layer. A thinner depletion layer produces a smaller electric field in the layer and a smaller potential difference across the layer.

Now that the potential difference is smaller across the depletion layer, it is significantly more likely that an electron in the n-region or a hole in the p-region will happen to have enough kinetic energy to move through the depletion layer against this (smaller) potential difference.

> **QUESTION** As a result of the decreased potential difference, what will happen to I_{tg}? To I_r?

The thermal-generation current I_{tg} won't change. It is made up of electrons and holes that are swept quickly through the depletion layer by the electric field anyway, and they will continue to be swept through quickly. A small initial kinetic energy $K_i \approx 0$ is sufficient.

The recombination current I_r, however, is made up of electrons and holes that had to travel against the electric force, and now it is easier to go against the smaller forces corresponding to the smaller potential difference. A large initial kinetic energy $K_i \approx E_{gap}$ is needed. Therefore the recombination current will increase rapidly with small decreases in potential difference across the depletion layer and may be very large compared to the thermal-generation current.

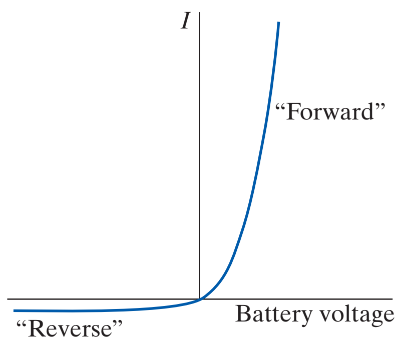


Figure S2.38 Current through a diode as a function of applied battery voltage. The battery connections are switched between the left and right sides of this graph.

> **QUESTION** What is the direction of the net current, which is approximately I_r, since it is now much larger than I_{tg}?

The recombination conventional current flows in the direction opposite to the electric field that is in the depletion layer, so it flows to the right, from the p-type region to the n-type region, in the expected direction for the way the battery is attached.

It is important to understand that this current is comparatively large because there are so many more mobile electrons and holes than in pure silicon. This large current increases rapidly with decreasing potential difference across the p-n junction (corresponding to increasing emf of the attached battery or power supply). In fact, the current increases exponentially rapidly. (In terms of the Boltzmann factor, the recombination current increases by an amount $e^{+e|\Delta V_{batt}|/k_B T}$, where ΔV_{batt} is the potential difference due to the battery.)

Figure S2.38 shows the net current through the diode for reverse and forward voltages. You see the small, nearly constant current for reverse voltages and the rapidly rising current for forward voltages. (Note that you have to switch the direction of the battery to make measurements corresponding to the left and right sides of this graph.)

A comment is in order about the detailed mechanism for the current in the forward direction in Figure S2.37. On the right, electrons flow into the diode from the battery and stream to the left across the p-n junction thanks to the lowered potential difference there. When they reach the bulk p-region, they recombine with the abundant holes. Electrons in the n-region are continually replenished by electrons coming from the battery.

Holes in the p-region stream to the right across the p-n junction, and when they reach the bulk n-region they recombine with the abundant electrons. It is difficult without a full quantum-mechanical analysis to explain in a satisfying way what replenishes the supply of holes. Briefly, at the junction of the metal wire with the p-region there is another depletion region, and with the electric field driving holes to the right, mobile electrons are created in the p-region near the junction and move into the metal wire, thereby replenishing the supply of holes near the junction.

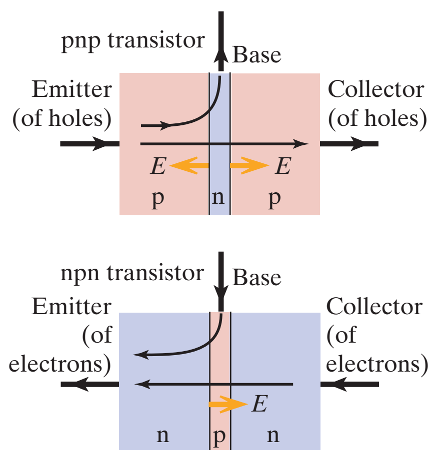


Figure S2.39 Bipolar transistors, pnp and npn. Direction of conventional current is shown. Also shown are the large electric fields in the depletion layers.

The Bipolar Transistor

One important kind of transistor, the bipolar transistor, consists of two back-to-back p-n junctions as shown in Figure S2.39. Both pnp and npn configurations are used. These devices have the property that a small current running into or out of the side of the thin middle region (called the base) has a big effect on how much current runs through the transistor between "emitter" and "collector." This makes possible current amplification, where a large collector current is proportional to a small base current. In another use the transistor can act as a kind of gate: depending on the base current, the collector current is on (large) or off (very small).

We will very briefly describe the basic operation of a pnp transistor. Voltages are applied to the pnp transistor at the top of Figure S2.39 so that hole current runs from the emitter in the "forward" or large-current direction. Small increases in emitter-base voltage lead to large increases in emitter current. The n-type region is made very thin, so that many holes manage to cross the thin n-type region. The large electric field in the depletion layer between the n-type region and the collector points to the right, so it sweeps to the right any holes that reach that point. The base current is quite small compared to the collector current.

With this arrangement, circuit elements attached to the collector are pretty much decoupled from the emitter-base part of the circuit. The large collector current can run through a resistor connected between collector and base and produce a large voltage. The net effect is that a small emitter-base voltage change can make a big change in the collector-base voltage, and the transistor acts as an amplifier. The analysis of an npn transistor, shown at the bottom of Figure S2.39, is similar but in terms of electrons rather than holes.

This description of bipolar transistors is far from complete, but we hope it will give you a basic idea of the functioning of a transistor. A good understanding of p-n junctions provides a useful foundation for further study.

We should mention that although the bipolar transistor was one of the first commonly used types of transistor, in large integrated circuits containing thousands of transistors these are typically "field effect transistors" (FETs). In a field effect transistor the flow of current is controlled not by a small current as in the bipolar resistor but by an applied electric field that increases or decreases the thickness of a conducting region.

An Alternative Calculation Method

We mention another calculational technique that is often used in semiconductor physics. Consider again our calculation using Gauss's law for the electric field in the depletion layer of a p-n junction, but instead of a

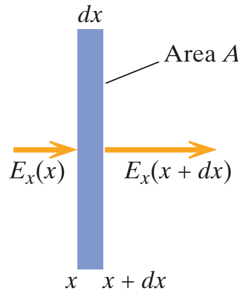


Figure S2.40 Applying Gauss's law to a box whose length is only dx.

long Gaussian box let the box have a very short length dx (Figure S2.40). The electric field at the left edge of the box is $E_x(x)$, and the electric field at the right edge is $E_x(x+dx)$. The area of the end of the box is again A as before, so the volume is Adx. Applying Gauss's law to this box in a region where the charge density is ρ gives this (the dielectric constant is K):

$$E_x(x+dx)A - E_x(x)A = \frac{Q}{K\varepsilon_0} = \frac{\rho A dx}{K\varepsilon_0}$$

$$\frac{E_x(x+dx) - E_x(x)}{dx} = \frac{\rho}{K\varepsilon_0}$$

However, $[E_x(x+dx) - E_x(x)]/dx$ in the limit of small dx is dE_x/dx, so

$$\frac{dE_x}{dx} = \frac{\rho}{K\varepsilon_0}$$

We could have started with this equation (Poisson's equation) instead of Gauss's law to determine the magnitude of the electric field in the depletion layer. For example, in the left half of the depletion region where the charge density is $-\rho$, we have

$$\frac{dE_x}{dx} = \frac{-\rho}{K\varepsilon_0}$$

Solving for E_x, we have the following, where the minus sign expresses the fact that the electric field points to the left in the depletion layer.

$$E_x = \frac{-\rho}{K\varepsilon_0}x$$

This is the same result we obtained earlier by a different use of Gauss's law.

SUMMARY

The quantum energy levels in crystalline solids occur in bands. In semiconductors there is a small energy gap between a completely filled "valence" band and an empty "conduction" band. The semiconductor becomes a conductor if electrons are promoted from the valence band to the conduction band, due to thermal effects or as a result of "doping" the semiconductor.

In a junction of p-type and n-type semiconductors there arises a "depletion layer" with an associated locked-in electric field. This makes possible diode action, in which current can flow easily in one direction but not in the other direction.

PROBLEM

••P1 The band-gap energy in germanium has been measured to be about 0.7 eV. A typical doping level in germanium is 1×10^{18} atoms per cm^3. The density of germanium is 5.4 grams per cm^3, and there are 73 grams per mole. The dielectric constant is 16. Calculate the width of the depletion layer in a germanium p-n junction. Also calculate the maximum electric field in the junction.