Chapter 1

Conduction in Semiconductors

1.1 Introduction

All solid-state devices, e.g. diodes and transistors, are fabricated from materials known as semiconductors. In order to understand the operation of these devices, the basic mechanism of how currents flow in them must be understood. This chapter covers the fundamentals of conduction in semiconductors. The chapter is not intended to be an extensive introduction to the area of solid-state physics. Instead, only those topics which lead to a better understanding of the macroscopic properties of semiconductors are covered. The mechanisms of conduction in a metal and in a semiconductor are compared. The effects of impurities on conduction in semiconductors are discussed. The formation of a p-n semiconductor junction is described and its conduction properties are discussed.

1.2 Classification of Conductors

Figure 1.1 illustrates a two-dimensional view of an atom that is called the Bohr model of the atom. It consists of a positively charged nucleus and a system of negatively charged electrons which rotates around the nucleus. In a neutral atom, the total charge is zero. This means that the positive charge on the nucleus is equal to the total negative charge on the electrons. The electrons are bound to the nucleus by the forces of attraction between oppositely charged particles. They are arranged systematically in layers called shells. The closer a shell is to the nucleus, the more tightly bound are the electrons in that shell to the atom. The shell closest to the nucleus can contain no more than two electrons. The outer shell can have no more than eight. The number in the shells in between is determined by the laws of quantum mechanics.

The outermost shell in an atom contains what are called the *valence electrons*. These govern the nature of chemical reactions of the elements. In addition, they play a large part in determining the electrical behavior of the elements and the crystalline structure of solids. The metallic elements tend to have one, two, or three valence electrons. The nonmetals have five, six, or seven. The inert gases have eight. The class of elements which have four valence electrons is called semiconductors. If a valence electron escapes its parent atom, it becomes free to move about. The parent atom then has a net positive charge and is called an *ionized atom* or an *ion*. If an electric field is applied to a material, the free electrons have forces exerted on them which cause them to move. This constitutes the flow of a current in the material that is called a *conduction current* or a *drift current*.

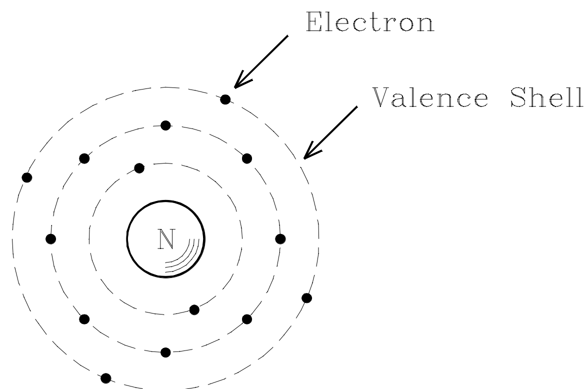

Figure 1.1: Two-dimensional Bohr model of an atom showing the nucleus and three shells.

Depending on the number n of free electrons per unit volume in a solid, the material is classified as being a good conductor, a semiconductor, or an insulator. For a good conductor, n is very large and is independent of temperature. A typical value is $n \simeq 10^{28}$ per m^3. For an insulator at ordinary temperatures, n is much smaller and has a typical value $n \simeq 10^7$ per m^3. For a semiconductor, it lies between the values for a good conductor and an insulator and is a function of the temperature. Silicon is an important semiconductor for which $n \simeq 1.5 \times 10^{16}$ per m^3 at room temperature ($T = 300\,\mathrm{K}$).

1.3 Conduction in Metals

Metals are classified as good conductors. The valence electrons are so loosely bound to the atoms that they are free to move about in the conductor. Fig. 1.2 shows a two-dimensional illustration of the atoms in a metal with the free electrons distributed randomly among the immobile ions. The free electrons can be visualized as molecules of a gas that permeate the region between the ions. Analogous to the random motion of molecules in a gas, thermal energy causes the free electrons to be in continuous random motion. Observation of an individual electron would reveal that its direction of motion changes randomly after each collision with an ion. Because the direction of motion of each electron is random, the average number of electrons passing through any area per unit time is zero. Thus the average current flow in the metal is zero.

1.3.1 Drift Velocity

If an electric field $\overrightarrow{E}$ (V/m) is applied to a metal, an electrostatic force is exerted on the free electrons which causes a conduction current to flow. (The arrow indicates a vector quantity.) The force on an individual electron is given by $\overrightarrow{F} = -q\overrightarrow{E}$ (N), where q is the electronic charge ($q = 1.602 \times 10^{-19}\,\mathrm{C}$). The electrostatic forces cause the electrons to be accelerated in a direction opposite to that of the applied field. Fig. 1.3 illustrates the path that an individual electron might take under the influence of the electric field. If the electron did not collide with the bound ions, its velocity would increase indefinitely. However, energy is lost with each collision so that the average

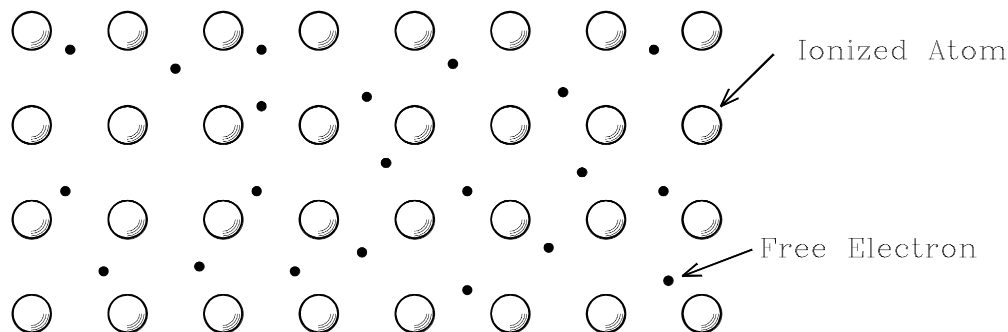

Figure 1.2: Two-dimensional view of the atoms in a metal with free electrons distributed randomly among the ions.

velocity approaches a constant or steady-state value. The average velocity $\overrightarrow{v}$ (m/s) is called the *drift velocity*. It is proportional to the applied field and is given by

$$\overrightarrow{v} = -\mu_e \overrightarrow{E} \tag{1.1}$$

where μ_e ($\mathrm{m^2\,V^{-1}\,s^{-1}}$) is the *electron mobility*. (The minus sign is required because the negative charge on the electron causes it to move is a direction opposite to the field.) The average distance that the electron travels between collisions with the bound ions is called the *mean free path*. As the temperature increases, the bound ions vibrate with increasing intensity, causing the mean free path between collisions to decrease. This effect causes the drift velocity $\overrightarrow{v}$ to decrease, which is modeled by a decrease in the electron mobility μ_e with temperature.

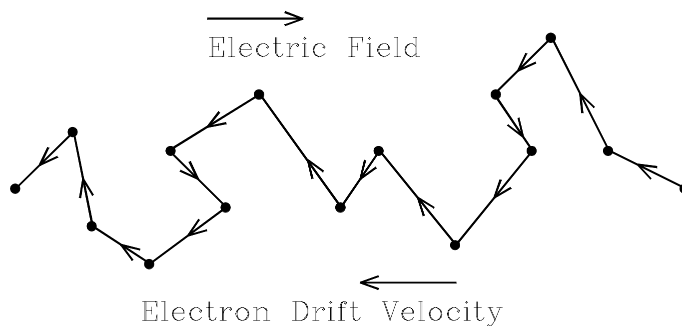

Figure 1.3: Path taken by a free electron in a metal under the influence of an applied electric field.

1.3.2 Charge Density

The *charge density* ρ ($\mathrm{C/m^3}$) in a conductor is defined as the free charge per unit volume. To relate the charge density in a metal to the density of free electrons, let n be the number of electrons

per m^3. Because the charge per electron is $-q$, it follows that the free charge per unit volume in the metal is given by

$$\rho = -nq \tag{1.2}$$

1.3.3 Current Density

The *current density* $\overrightarrow{J}$ ($\mathrm{A/m^2}$) in a conductor is defined as the current per unit area flowing in a particular direction. To relate the current density in a conductor to the drift velocity of the moving charges, consider a section of wire of length $\Delta\ell$ in which a current I is flowing. This is illustrated in Fig. 1.4. The charge in the section is $\Delta Q = \rho\Delta V = \rho S\Delta\ell$, where ρ is the charge density and S is the cross-sectional area of the wire. Let Δt be the time required for the charge in the section to move the distance $\Delta\ell$. The velocity of the charge is $\overrightarrow{v} = \widehat{a}\Delta\ell/\Delta t$, where $\widehat{a}$ is a unit vector in the direction of current flow. The current I flowing in the area S is $I = \Delta Q/\Delta t$. It follows that the current density $\overrightarrow{J}$ can be related to the drift velocity $\overrightarrow{v}$ as follows:

$$\overrightarrow{J} = \widehat{a}\frac{I}{S} = \widehat{a}\frac{\Delta Q}{S\Delta t} = \widehat{a}\frac{\rho S\Delta\ell}{S\Delta t} = \rho\left(\widehat{a}\frac{\Delta\ell}{\Delta t}\right) = \rho\overrightarrow{v} \tag{1.3}$$

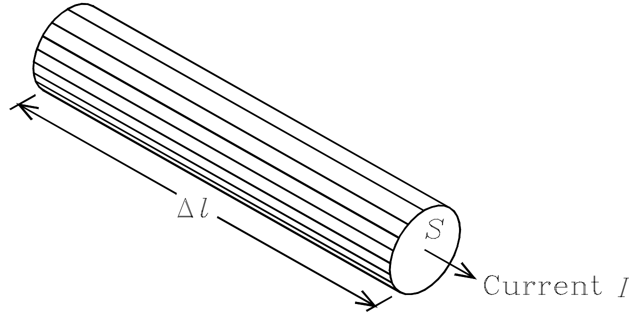

Figure 1.4: Section of wire of length $\Delta\ell$ in which a current I flows.

1.3.4 Conductivity

Using Eqs. (1.1) through (1.3), we can relate the current density $\overrightarrow{J}$ to the electric field $\overrightarrow{E}$ in a metal as follows:

$$\overrightarrow{J} = \rho\overrightarrow{v} = (-nq)\left(-\mu_e\overrightarrow{E}\right) = nq\mu_e\overrightarrow{E} = \sigma\overrightarrow{E} \tag{1.4}$$

(Note that two minus signs have canceled so that $\overrightarrow{J}$ is in the direction of $\overrightarrow{E}$.) This equation defines the *conductivity* σ ($\Omega^{-1}\,\mathrm{m}^{-1}$) of the metal. It is given by

$$\sigma = nq\mu_e \tag{1.5}$$

Because n is independent of temperature in a metal, it follows that the decrease in electron mobility μ_e with temperature causes the conductivity σ to decrease with temperature.

Example 1 *Aluminum has three valence electrons per atom, an atomic weight of* $0.02698\,\mathrm{kg/\,mol}$*, a density of* $2700\,\mathrm{kg/\,m^3}$*, and a conductivity of* $3.54\times10^7\,\mathrm{S/\,m^{-1}}$*. Calculate the electron mobility in aluminum. Assume that all three valence electrons in each atom are free.*

Solution. Recall from introductory chemistry that a mole of any substance is a quantity equal to its atomic weight and contains a number of molecules equal to Avogadro's number which is 6.02×10^{23}. It follows that the number of aluminum atoms per $\mathrm{m^3}$ is

$$6.02\times10^{23}\frac{\text{atoms}}{\text{mol}}\times\frac{1}{0.02698}\frac{\text{mol}}{\text{kg}}\times2700\frac{\text{kg}}{\text{m}^3}=6.024\times10^{28}\frac{\text{atoms}}{\text{m}^3}$$

Thus the electron density in the aluminum is $n=3\times6.024\times10^{28}=1.807\times10^{29}$ per $\mathrm{m^3}$. From Eq. (1.5), the mobility is given by

$$\mu_e=\frac{\sigma}{nq}=\frac{3.54\times10^7}{6.024\times10^{28}\times1.602\times10^{-19}}=3.67\times10^{-3}\,\mathrm{m^2\,V^{-1}\,s^{-1}}$$

1.3.5 Resistance

Consider the section of wire illustrated in Fig. 1.4. A conduction current I is flowing in the wire so that the current density is $\overrightarrow{J}=\widehat{a}I/S$, where $\widehat{a}$ is a unit vector in the direction of current flow and S is the cross-sectional area. Let V be the voltage drop across the section of length $\Delta\ell$ so that the electric field in the section is $\overrightarrow{E}=\widehat{a}V/\Delta\ell$. With the aid of Eq. (1.4), we can write $\overrightarrow{J}=\widehat{a}\sigma V/\Delta\ell$. By equating the two relations for $\overrightarrow{J}$, we obtain

$$\overrightarrow{J}=\widehat{a}\frac{I}{S}=\widehat{a}\frac{\sigma V}{\Delta\ell} \tag{1.6}$$

This equation can be solved for the resistance R of the section of wire to obtain

$$R=\frac{V}{I}=\frac{\Delta\ell}{\sigma S} \tag{1.7}$$

Thus the resistance is directly proportional to the length of the wire and inversely proportional to its area. Because the conductivity σ decreases with temperature, it follows from this equation that R increases with temperature. In most metals, the resistance increases linearly with temperature.

Example 2 *The conductivity of copper is* $5.8\times10^7\,\mathrm{S/\,m}$*. If a* $1\,\mathrm{m}$ *length of copper wire has a resistance of* $1\,\Omega$*, what is the thickness of the wire? Assume a circular cross section.*

Solution. Let d be the diameter of the wire. Using Eq. (1.7), we can write $S=\pi\,(d/2)^2=\Delta\ell/\sigma R=1/\left(5.8\times10^7\times1\right)$. Solution for d yields $d=0.148\,\mathrm{mm}$.

1.4 Conduction in Intrinsic Semiconductors

Semiconductors are the class of elements which have four valence electrons. Two important semiconductors are germanium (Ge) and silicon (Si). Early solid-state electronic devices were fabricated almost exclusively from germanium, whereas modern devices are fabricated almost exclusively from silicon. Gallium arsenide (GaAs) is a semiconductor compound made up of gallium, which has three

valence electrons, and arsenic, which has five. This semiconductor is making inroads in digital applications which require extremely high switching speeds and in extremely high-frequency analog applications. However, silicon remains the most useful semiconductor material and is expected to dominate for many years to come.

Semiconductor materials are normally in crystalline form with each valence electron shared by two atoms. The semiconductor is said to be *intrinsic* if it is not contaminated with impurity atoms. Fig. 1.5 shows a two-dimensional view of an intrinsic semiconductor crystal. Each circle represents both the nucleus of an atom and all electrons in that atom except the valence electrons. The links between the circles represent the valence electrons. Each valence electron can be assumed to spend half time with each of two atoms so that each atom sees eight half-time electrons. Compared to a metal, the valence electrons in a semiconductor are tightly bound.

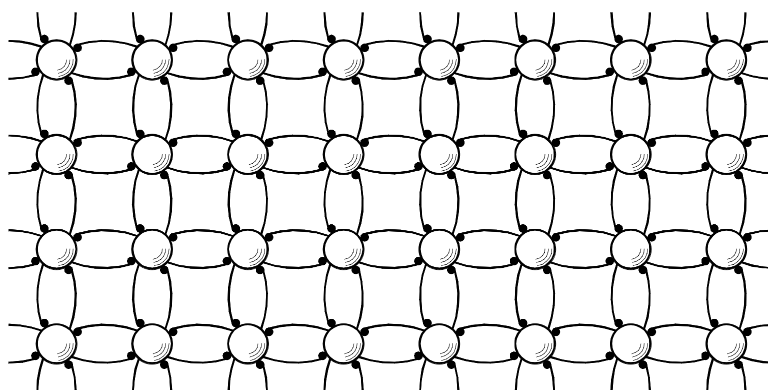

Figure 1.5: Two-dimensional illustration of the crystal lattice of an intrinsic semiconductor.

The thermal energy stored in a semiconductor crystal lattice causes the atoms to be in constant mechanical vibration. At room temperature, the vibrations shake loose several valence electrons which then become free electrons. In intrinsic silicon, the number of free electrons is approximately one in 10^{12} of the total number of valence electrons. The free electrons behave similarly to those in a metal. Under the influence of an applied electric field, they have a mobility and exhibit a drift velocity which produces a conduction current. However, because of the small number of free electrons, the conductivity of an intrinsic semiconductor is much lower than that of a metal.

When an electron is shaken loose from an atom, an electron vacancy is left which is called a *hole*. The parent atom then becomes an ion. The constant mechanical vibration of the lattice can cause the ion to capture a valence electron from a neighboring atom to replace the missing one. When such a transfer takes place, the position of the hole moves from one atom to another. This is equivalent to a positive charge $+q$ moving about in the semiconductor. (The motion of a hole can be likened to the motion of a bubble in water.) Like free electrons, holes have a mobility and exhibit a drift velocity which produces a conduction current under the influence of an applied electric field. Because of the opposite charge polarity of electrons and holes, they drift in opposite directions under the influence of a field.

Figure 1.6 illustrates the drift of free electrons in an intrinsic semiconductor under the application of an electric field that is directed from left to right. When an electron is shaken loose from its valence shell, an electron-hole pair is formed. The force generated by the electric field causes the

free electrons to drift to the left. Fig. 1.7 illustrates the drift of holes. In effect, a hole drifts to the right when a bound valence electron shifts to the left from one atom to another. The arrows in the figure point from the new position of a hole to its former position, i.e. in the direction of movement of the bound electrons in the lattice. The movement of holes may be likened to the movement of bubbles of air in water, where the water represents the bound electrons and the bubbles represent the holes. The movement of a bubble in one direction is really the result of a movement of water in the opposite direction. In summary, the flow of current in the semiconductor is the result of the flow of two components. One component is the flow of free electrons in one direction. The other component is the flow of the absence of bound electrons in the other direction. Because of the opposite charge polarities, the electron current and the hole current add to produce the total conduction current.

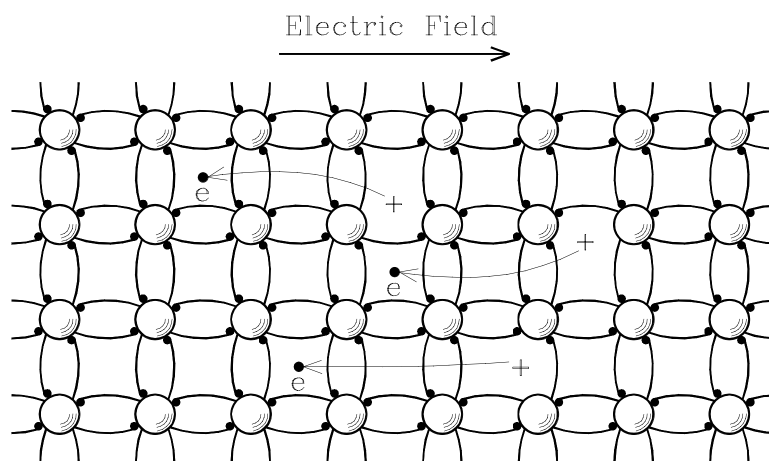

Figure 1.6: Illustration of the drift of free electrons under the application of an external electric field.

1.4.1 Recombinations

Because hole-electron pairs are continually created by thermal agitation of a semiconductor lattice, it might seem that the number of holes and free electrons would continually increase with time. This does not happen because free electrons are continually recombining with holes. At any temperature, a stable state is reached when the creation rate of hole-electron pairs is equal to the recombination rate. The *mean lifetime* τ_n (s) of a free electron is the average time that the electron exists in the free state before recombination. The mean lifetime τ_p (s) for the hole is defined similarly. In the intrinsic semiconductor, τ_n is equal to τ_p because the number of free electrons must be equal to the number of holes. However, the addition of an impurity to the semiconductor lattice can cause the mean lifetimes to be unequal.

1.4.2 Intrinsic Concentration

Denote the number of free electrons per m^3 in a semiconductor by n and the number of holes per m^3 by p. In an intrinsic semiconductor, the hole concentration must equal the electron concentration.

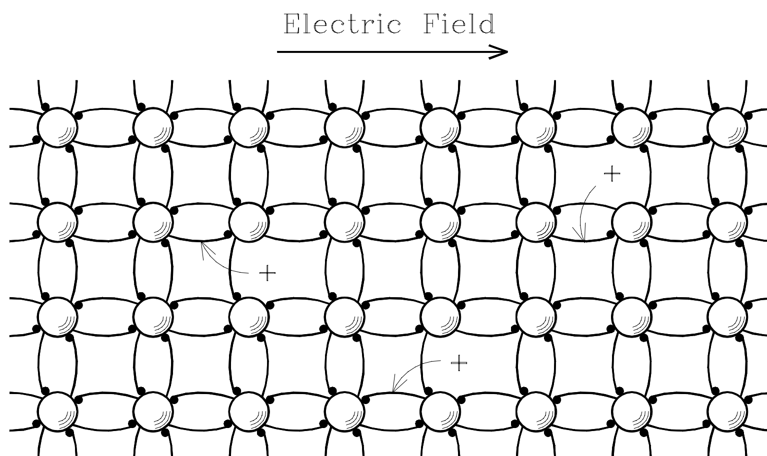

Figure 1.7: Illustration of the drift of holes under the application of an external electric field.

In this case we write $n = p = n_i$, where n_i is called the *intrinsic concentration.* It can be shown that n_i can be written

$$n_i = n_0 T^{3/2} \exp\left(\frac{-V_G}{2V_T}\right) \tag{1.8}$$

where n_0 is a constant, T is the absolute temperature, V_G is the semiconductor bandgap voltage, and V_T is the thermal voltage. (The bandgap voltage multiplied by q represents the minimum energy required to cause a bound valence electron to become a free electron.) For silicon, the bandgap voltage at $T = 300\,\mathrm{K}$ has the value $V_G = 1.11\,\mathrm{V}$. The thermal voltage is related to the temperature by

$$V_T = \frac{kT}{q} \tag{1.9}$$

where k is the Boltzmann constant ($k = 1.381 \times 10^{-23}\,\mathrm{J/K}$). At $T = 300\,\mathrm{K}$, the thermal voltage has the value $V_T = 0.0259\,\mathrm{V}$.

Example 3 *The initial temperature of a specimen of silicon is* $T_1 = 300\,\mathrm{K}$. *By what factor does the intrinsic concentration* n_i *increase if the temperature increases by* $10\,^\circ\mathrm{C}$ $(18\,^\circ\mathrm{F})$*? Assume the bandgap voltage at both temperatures is* $V_G = 1.11\,\mathrm{V}$.

Solution. Let $T_1 = 300\,\mathrm{K}$ be the initial temperature and $T_2 = 310\,\mathrm{K}$ the final temperature. At $300\,\mathrm{K}$, the thermal voltage is $V_{T1} = 0.0259\,\mathrm{V}$. At $310\,\mathrm{K}$, it is $V_{T2} = (1.380 \times 10^{-23} \times 310)/1.602 \times 10^{-19} = 0.0267\,\mathrm{V}$. The factor by which n_i increases is calculated from Eq. (1.8) as follows:

$$\frac{n_{i2}}{n_{i1}} = \frac{(T_2)^{3/2} \exp\left(-V_G/2V_{T2}\right)}{(T_1)^{3/2} \exp\left(-V_G/2V_{T1}\right)} = \left(\frac{310}{300}\right)^{3/2} \exp\left[\frac{1.11}{2}\left(\frac{-1}{0.0267} + \frac{1}{0.0259}\right)\right] = 2.0$$

1.4.3 Conductivity

When an electric field $\overrightarrow{E}$ is applied to an intrinsic semiconductor, the free electrons drift with a velocity $\overrightarrow{v}_e = -\mu_e \overrightarrow{E}$, where μ_e is the electron mobility. The holes drift with a velocity $\overrightarrow{v}_h = +\mu_h \overrightarrow{E}$,

where μ_h is the hole mobility. Although the free electrons and holes drift in opposite directions, the current densities add because the charge polarities are opposite. The electron charge density is $\rho_e = -n_i q$ and the hole charge density is $\rho_h = +n_i q$, where n_i is the intrinsic concentration. The total conduction current density can be written

$$\overrightarrow{J} = \rho_e \overrightarrow{v}_e + \rho_h \overrightarrow{v}_h = n_i (\mu_e + \mu_h) q \overrightarrow{E} = \sigma \overrightarrow{E} \tag{1.10}$$

This equation defines the conductivity σ of the intrinsic semiconductor. It is given by

$$\sigma = n_i (\mu_e + \mu_h) q \tag{1.11}$$

Example 4 *A rod of intrinsic silicon is* $1\,\mathrm{cm}$ *long and has a diameter of* $1\,\mathrm{mm}$. *At room temperature, the intrinsic concentration in the silicon is* $n_i = 1.5 \times 10^{16}$ *per* m^3. *The electron and hole mobilities are* $\mu_e = 0.13\,\mathrm{m}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$ *and* $\mu_h = 0.05\,\mathrm{m}^2\,\mathrm{V}^{-1}\,\mathrm{s}^{-1}$. *Calculate the conductivity* σ *of the silicon and the resistance* R *of the rod.*

Solution. The conductivity is calculated from Eq. (1.11) as follows:

$$\begin{aligned} \sigma &= n_i (\mu_e + \mu_h) q = 1.5 \times 10^{16} \times (0.13 + 0.05) \times 1.602 \times 10^{-19} \\ &= 4.33 \times 10^{-4}\,\mathrm{S/\,m} \end{aligned}$$

The resistance is calculated from Eq. (1.5) as follows:

$$R = \frac{\Delta \ell}{\sigma S} = \frac{0.01}{4.33 \times 10^{-4} \times \pi \left(0.5 \times 10^{-3}\right)^2} = 29.4\,\mathrm{M}\Omega$$

1.5 n-Type and p-Type Semiconductors

The preceding example illustrates how poor a conductor intrinsic silicon is at room temperature. The conductivity can be increased by adding certain impurities in carefully controlled minute quantities. When this is done, the semiconductor is called a *doped semiconductor*. There are two classes of impurities that are used. These are *donor impurities* and *acceptor impurities*. Typically one impurity atom is added per 10^8 semiconductor atoms. A semiconductor that is doped with a donor impurity is called an n-type *semiconductor*. One that is doped with an acceptor impurity is called a p-*type semiconductor*.

1.5.1 n-Type Semiconductor

An n-type semiconductor is produced by adding a donor impurity such as arsenic, antimony, or phosphorus to an intrinsic semiconductor. Each donor atom has five valence electrons. When a donor atom replaces an atom in the crystal lattice, only four valence electrons are shared with the surrounding atoms. The fifth valence electron becomes a free electron as illustrated in Fig. 1.8. The number of free electrons donated by the donor atoms is much greater than the number of free electrons and holes in the intrinsic semiconductor. This makes the conductivity of the n-type semiconductor much greater that of the intrinsic semiconductor. Because the number of free electrons is far greater than the number of holes, the free electrons are the *majority carriers*. The semiconductor is called n-type because the majority carriers have a negative charge.

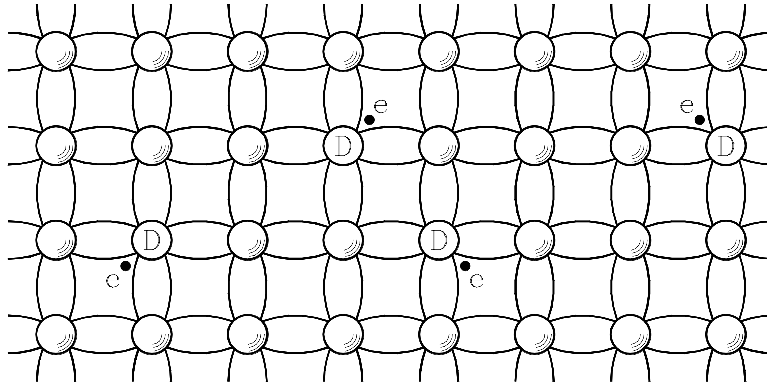

Figure 1.8: Two-dimensional illustration of the crystal lattice of an n-type semiconductor

Hole-electron pairs are continually formed by thermal agitation of the lattice in an n-type semiconductor. Because of the large number of donor electrons, there are many more free electrons available for recombination with the holes. This decreases the mean lifetime for the holes which decreases the number of holes in the n-type semiconductor compared to the intrinsic semiconductor. For this reason, the current due to the flow of holes in an n-type semiconductor is often neglected in calculations.

It is important to understand that a donor atom is electrically neutral if its fifth valence electron does not become a free electron in the lattice. If the fifth electron becomes a free electron, the number of protons in the atom is greater than the number of electrons by one. In this case, the donor atom becomes a bound positively charged ion.

1.5.2 p-Type Semiconductor

A p-type semiconductor is produced by adding an acceptor impurity such as gallium, boron, or indium to an intrinsic semiconductor. Each acceptor atom has three valence electrons. When an acceptor atom replaces an atom in the crystal lattice, there are only three valence electrons shared with the surrounding atoms. This leaves a hole as illustrated in Fig. 1.9. The number of holes created by the acceptor atoms is much greater than the number of free electrons and holes in the intrinsic semiconductor. This makes the conductivity of the p-type semiconductor much greater that of the intrinsic semiconductor. Because the number of holes is far greater than the number of electrons, the holes are the majority carriers. The semiconductor is called p-type because the majority carriers have a positive charge.

Hole-electron pairs are continually formed by thermal agitation of the lattice in a p-type semiconductor. Because of the large number of holes, there are many more holes available for recombination with the free electrons. This decreases the mean lifetime for the free electrons which decreases the number of electrons in the p-type semiconductor compared to the intrinsic semiconductor. For this reason, the current due to the flow of free electrons in a p-type semiconductor is often neglected in calculations.

It is important to understand that an acceptor atom is electrically neutral if the hole created by the absence of its fourth valence electron is not filled by an electron from an adjacent silicon atom.

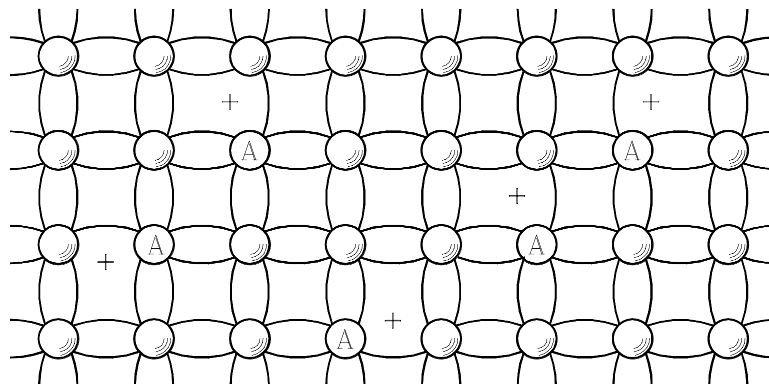

Figure 1.9: Two-dimensional illustration of the crystal lattice of a p-type semiconductor.

Once an electron fills the hole, the number of electrons in that atom is greater than the number of protons by one. In this case, the acceptor atom becomes a bound negatively charged ion.

1.5.3 Mass-Action Law

In an intrinsic semiconductor, we have noted that the electron concentration and the hole concentration are both equal to the intrinsic concentration, *i.e.* $n = p = n_i$. If this were not true, the material would not be electrically neutral. We have seen that adding an n-type impurity to the semiconductor increases n and decreases p. Similarly, adding a p-type impurity increases p and decreases n. It can be shown that the product of n times p is a constant independent of the doping type and the doping level. The product is given by

$$np = n_i^2 \tag{1.12}$$

where n_i is given by Eq. (1.8). This relation is called the *mass-action law*.

To understand this equation, consider an intrinsic semiconductor in which $n = p = n_i$. Assume that donors with the density $D_A = n_i$ are added to the semiconductor at $t = 0$. This initially doubles the total number of free electrons, which causes the recombination rate with the holes to double. This causes the hole density to drop from n_i to $n_i/2$. The initial free electron density is $2n_i$ which drops to $2n_i - n_i/2$ after the increase in recombinations. Thus the product of the electron and hole concentrations is $(2n_i - n_i/2) \times n_i/2 = 3n_i^2/4$. Now suppose that the number of donors is increased by the factor N, where N is large. In this case, the product becomes $(Nn_i - n_i/N) \times n_i/N = n_i^2\left(1 - 1/N^2\right) \simeq n_i^2$. Although this is not an exact proof, it illustrates the basic mechanism. A more detailed proof requires an involved thermodynamic analysis.

1.5.4 Electrical Neutrality

An intrinsic semiconductor is electrically neutral, i.e. there is no net charge stored. The addition of n-type or p-type impurities does not change this. To state this mathematically, let N_D be the number of donor atoms per m^3 and N_A the number of acceptor atoms per m^3. We assume that all donor atoms and all acceptor atoms are ionized so that there are N_D bound positive charges per

m^3 and N_A bound negative charges per m^3. Each donor ion has a charge $+q$ and each acceptor ion has a charge $-q$. The total number of negative charges per m^3 is equal to the number n of free electrons per m^3 plus the number N_A of bound acceptor atoms per m^3, i.e. $n + N_A$. Similarly, the number of positive charges per m^3 is equal to the number p of holes per m^3 plus the number N_D of bound donor atoms per m^3, i.e. $p + N_D$. Because the semiconductor is electrically neutral, the number of positive charges must equal the number of negative charges. This gives the condition

$$n + N_A = p + N_D \tag{1.13}$$

In an n-type semiconductor, $N_A = 0$ and $p << n$ so that the above equation can be solved for n to obtain

$$n = p + N_D \simeq N_D \tag{1.14}$$

The approximation in this equation and Eq. (1.12) can be used to solve for the hole concentration p to obtain

$$p \simeq \frac{n_i^2}{N_D} \tag{1.15}$$

Similarly, in a p-type semiconductor, we can write

$$p = n + N_A \simeq N_A \tag{1.16}$$

$$n \simeq \frac{n_i^2}{N_A} \tag{1.17}$$

Example 5 *In the silicon rod of Example 4, the number of silicon atoms per* m^3 *is* 5×10^{28}*. A donor impurity is added to the silicon in the concentration of one donor atom per* 10^8 *atoms of silicon. Calculate the new resistance of the rod. Assume that each donor atom contributes one free electron.*

Solution. The donor concentration in the silicon is calculated as follows:

$$N_D = 5 \times 10^{28} \frac{\text{atoms}}{\mathrm{m}^3} \times \frac{1}{1 \times 10^8} \frac{\text{donors}}{\text{atom}} = 5 \times 10^{20} \frac{\text{donors}}{\mathrm{m}^3}$$

It follows from Eq. (1.14) that the free electron concentration is $n \simeq N_D = 5 \times 10^{20}$ electrons per m^3. From Eq. (1.15), the hole concentration is $p \simeq n_i^2/n = (1.5 \times 10^{15})^2/5 \times 10^{20} = 4.5 \times 10^9$ holes per m^3. Because $p << n$, we can neglect p in calculating the conductivity. Eq. (1.11) gives $\sigma \simeq nq\mu_e = 5 \times 10^{20} \times 1.602 \times 10^{-19} \times 0.13 = 10.41\,\mathrm{S/m}$. The resistance is calculated from Eq. (1.5) as follows:

$$R = \frac{\Delta\ell}{\sigma S} = \frac{0.01}{10.41 \times \pi \left(0.5 \times 10^{-3}\right)^2} = 1.22\,\mathrm{k\Omega}$$

Compared to the intrinsic silicon rod of Example 2, this is smaller by a factor of 24,100.

It is possible to add both an acceptor impurity and a donor impurity to an intrinsic semiconductor. If the donor concentration N_D is equal to the acceptor concentration N_A, the semiconductor remains intrinsic because the free electrons of the donors combine with the holes of the acceptors. With $N_D = N_A$, Eqs. (1.12) and (1.13) give $n = p = n_i$. If $N_D > N_A$, the semiconductor becomes an n-type. In this case, Eqs. (1.14) and (1.15) become

$$n = p + N_D - N_A \simeq N_D - N_A \tag{1.18}$$

$$p \simeq \frac{n_i^2}{N_D - N_A} \tag{1.19}$$

Similarly if $N_A > N_D$, the semiconductor becomes a p-type and Eqs. (1.16) and (1.17) hold if N_A is replaced by $N_A - N_D$.

$$p = n + N_A - N_D \simeq N_A - N_D \tag{1.20}$$

$$n \simeq \frac{n_i^2}{N_A - N_D} \tag{1.21}$$

1.5.5 Conductivity

The conduction current density in an intrinsic semiconductor is given by Eq. (1.10). In a doped semiconductor, it is given by

$$\overrightarrow{J} = (n\mu_e + p\mu_h)\, q\overrightarrow{E} = \sigma \overrightarrow{E} \tag{1.22}$$

where n is the electron concentration and p is the hole concentration. This equation defines the conductivity σ. It is given by

$$\sigma = (n\mu_e + p\mu_h)\, q \tag{1.23}$$

For an intrinsic semiconductor, $n = p = n_i$ and this equation reduces to Eq. (1.11).

1.5.6 Diffusion Current

In an n-type or a p-type semiconductor, it is possible to have a component of current that is not a conduction current. This current is due to the non-uniform density of free electrons or holes and is called a *diffusion current*. It is not possible to have a diffusion current in a metal. To achieve a non-uniform density of free electrons or holes, the doping concentration in the semiconductor is not constant, i.e. it is a function of position. Such a concentration is called a *graded doping concentration*.

Figure 1.10 illustrates a semiconductor in which the concentration of holes is a function of the coordinate z, i.e. $p = p(z)$. In addition, $p(z)$ is a decreasing function of z. Consider the plane defined by $z = z_1$. Because the number of holes to the left of the plane is greater than the number to the right, it is reasonable to expect that the random motion of the holes due to thermal energy would cause more holes to migrate from left to right than from right to left. Thus there is a net current flow across the plane from left to right. This current is called a diffusion current.

In general, the hole diffusion current density is given by

$$\overrightarrow{J}_p = -qD_h \nabla p \tag{1.24}$$

where D_h is the hole diffusion constant and ∇p is the gradient or directional derivative of p. If $\widehat{x}$, $\widehat{y}$, and $\widehat{z}$, respectively, are unit vectors in the x, y, and z directions, ∇p is given by

$$\nabla p = \widehat{x}\frac{\partial p}{\partial x} + \widehat{y}\frac{\partial p}{\partial y} + \widehat{z}\frac{\partial p}{\partial z} \tag{1.25}$$

For the case illustrated in Fig. 1.10, p is a function of z only so that $\nabla p = \widehat{z}dp/dz$. (The total derivative is used because p is a function of z only.) Because $p(z)$ is a decreasing function of z in the figure, it follows that $dp/dz < 0$. This makes the direction of the diffusion current in the $+z$ direction.

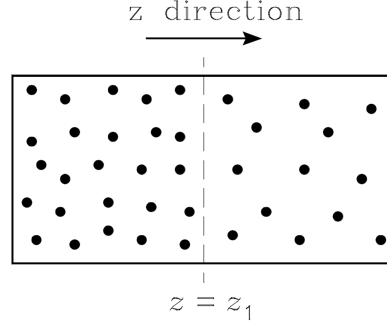

Figure 1.10: Illustration of a semiconductor in which the hole concentration is a function of the coordinate z.

In an n-type semiconductor that has a non-uniform density of free electrons, the electron diffusion current is given by

$$\overrightarrow{J}_n = qD_e\nabla n$$

1.25 where n is the electron concentration function and D_e is the electron diffusion constant. The diffusion constants are related to the mobility constants by the relation

$$\frac{D_e}{\mu_e} = \frac{D_h}{\mu_h} = V_T \tag{1.26}$$

where V_T is the thermal voltage given by Eq. (1.9). This relation is known as the *Einstein equation*.

1.5.7 Total Current

In general, the total current in a semiconductor is written as the sum of the electron and hole conduction currents and the electron and hole diffusion currents. It is given by

$$\overrightarrow{J} = \sigma\overrightarrow{E} + q(D_e\nabla n - D_h\nabla p) \tag{1.27}$$

where $\overrightarrow{E}$ is the electric field intensity and σ is the conductivity given by Eq. (1.23). In an open-circuited semiconductor, the equilibrium current density must be zero. If we set $\overrightarrow{J} = 0$, the above equation predicts an electric field in the open-circuited semiconductor given by

$$\overrightarrow{E} = \frac{q}{\sigma}\left(D_h\nabla p - D_e\nabla n\right) = \frac{V_T}{n\mu_e + p\mu_h}\left(\mu_h\nabla p - \mu_e\nabla n\right) \tag{1.28}$$

where Eqs. (1.23) and (1.26) have been used.

As an application of the preceding results, let us calculate the voltage difference between two points in a semiconductor caused by a graded doping concentration. Consider a p-type semiconductor in which the hole concentration is a function of the coordinate z, i.e. $p = p(z)$. We assume that there are no external sources connected to the semiconductor so that the equilibrium current density is zero. The semiconductor is illustrated in Fig. 1.11. Let the voltage at $z = z_1$ be V_1 and the voltage at $z = z_2$ be V_2. The hole concentrations are labeled p_1 and p_2. We assume that $p(z)$

is a decreasing function of z so that $p_1 > p_2$. This gives an electric field $\overrightarrow{E}$ that is directed in the negative z direction.

Because there are more holes at z_1 than at z_2, it may seem that the electric field should be directed from left to right. To understand why it is directed to the left, assume that $p_1 = p_2$ initially. In this case, there is no initial electric field. At $t = 0$, let the hole density at z_1 be increased by the addition of acceptors. As t increases, some of the holes at z_1 diffuse toward z_2, leaving bound negative ions at z_1. The diffusion of holes toward z_2 increases the net positive charge at z_2. Thus the voltage at V_2 becomes positive with respect to the voltage at V_1 and the electric field is directed from right to left. When equilibrium is reached, the net current flow is zero because the force generated by the electric field cancels the diffusion force.

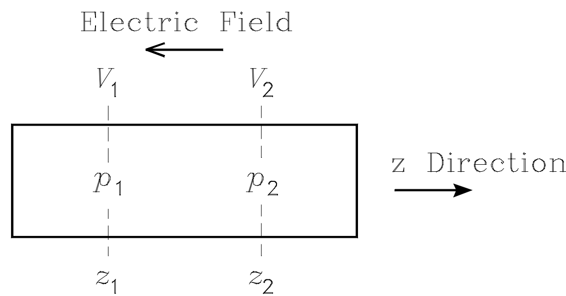

Figure 1.11: p-type semiconductor with nonuniform doping.

To solve for the electric field $\overrightarrow{E}$, we set $n = 0$ and $\nabla p = \widehat{a}_z dp/dz$ in Eq. (1.28) to obtain

$$\overrightarrow{E} = \widehat{a}_z \frac{V_T}{p} \frac{dp}{dz} \tag{1.29}$$

Because $\overrightarrow{E}$ is related to the voltage or potential function V by $\overrightarrow{E} = -\nabla V$ and $V = V(z)$, it follows that $\overrightarrow{E} = -\widehat{z} dV/dz$. By equating the two expressions for $\overrightarrow{E}$, we obtain

$$\widehat{z}\frac{V_T}{p} \frac{dp}{dz} = -\widehat{z}\frac{dV}{dz} \tag{1.30}$$

when the $\widehat{z}$ and the dz are canceled from both sides of the equation, we obtain

$$dV = V_T \frac{dp}{p}$$

This can be integrated to obtain

$$V_2 - V_1 = V_T \ln\left(\frac{p_1}{p_2}\right) \tag{1.31}$$

It follows that the voltage difference depends only on the concentrations at the two points and is independent of the separation of the points. A similar equation can be derived for the voltage difference as a function of the free electron concentration n in an n-type semiconductor. It is given by

$$V_2 - V_1 = -V_T \ln\left(\frac{n_1}{n_2}\right) \tag{1.32}$$

1.6 The Open-Circuited p-n Junction

A *p-n junction* is the junction between an n-type semiconductor and a p-type semiconductor. It is fabricated by introducing donor impurities into one side of an intrinsic semiconductor crystal and acceptor impurities into the other side. The transition between the two regions occurs in a very small distance, typically $0.5\,\mu$m. Fig. 1.12 illustrates the cross section of a p-n junction where the donor ions are represented by positive signs and the acceptor ions are represented by negative signs. Initially, we assume that the only charge carriers in the n-type side are free electrons and that the only charge carriers in the p-type side are holes.

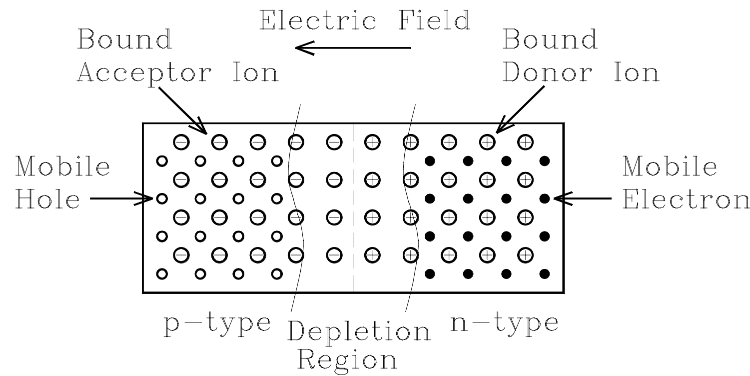

Figure 1.12: Diagram of a p-n junction with the width of the depletion region greatly exaggerated.

Because of the unequal electron concentrations and unequal hole concentrations on the two sides of the junction, a diffusion current consisting of both holes and free electrons will flow across the junction. (The diffusion process is similar to the diffusion of different gases between two glass jars joined at the mouths.) Holes diffuse out of the p-type side and into the n-type side and free electrons diffuse out of the n-type side and into the p-type side. This causes the p-type side to become negatively charged and the n-type side to become positively charged. The charges cause an electric field to build up across the junction which is directed from the n-type side to the p-type side. The polarity of the electric field is such that the force it exerts on the holes and free electrons opposes the diffusion process. Equilibrium is reached when the force exerted on the charge carriers by the electric field is equal to the diffusion force.

Let us now consider what happens when thermal agitation of the semiconductor lattice produces a hole-electron pair in the region near the junction. The electric field directed from the n-type side to the p-type side exerts a force on the free electron and causes it to be swept to the n-type side. Similarly, the hole is swept to the p-type side. The directions that the charges move are opposite to those due to the diffusion process. When equilibrium is reached, the net number of both electrons and holes crossing the junction is zero.

1.6.1 Depletion Region

Because no free electrons or holes can exist is the region about the junction, there are no mobile charges to neutralize the ions in this region. This is illustrated in Fig. 1.12. The ions on the

n-type side have a positive charge on them and those on the p-type side have a negative charge. These charges are called *uncovered charges*. The region about the junction in which the uncovered charges exist is called the *depletion region*. Other names for this are the *space-charge region* and the *transition region*. Fig. 1.13a illustrates the plot of the net uncovered charge density in the p-n junction as a function of distance from the junction. The charge distribution is called a dipole distribution because the charge on one side of the junction is the negative of the charge on the other side. The uncovered charges on each side of the junction can be thought of as the charges on the plates of a parallel plate capacitor as shown in Fig. 1.13b.

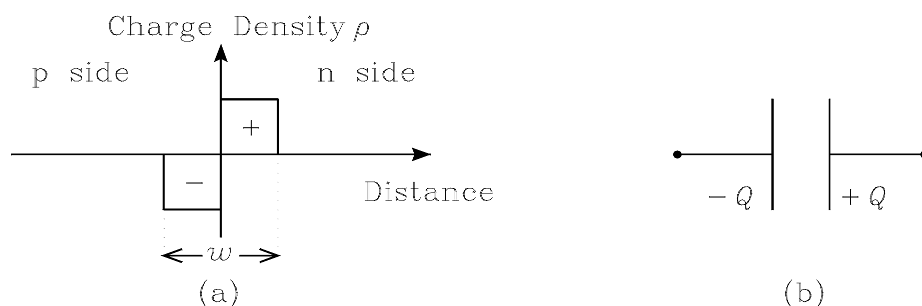

Figure 1.13: (a) Plot of the charge density as a function of distance from the junction. (b) Parallel plate capacitor analog of the charge distribution.

Because of charge neutrality, the total uncovered charge on the n-type side of the depletion region must be equal to the negative of the total uncovered charge on the p-type side. If the n and p concentrations are equal, it follows that the widths of the uncovered charge regions on the two sides of the junction must be equal. Now, suppose the p concentration is increased while holding the n concentration constant. Charge neutrality requires the width of the p-type side of the depletion region to decrease if the total uncovered charge is to remain constant. Similarly, if the n concentration is increased while holding the p concentration constant, the width of the n-type side must decrease. We conclude, in general, that increasing either p or n or both decreases the total width w of the depletion region illustrated in Fig. 1.13a. This has an important effect on the reverse-bias breakdown characteristics of a junction. This is discussed in the following chapter.

1.6.2 Built-In Potential

Because there is an electric field in the depletion region of a p-n junction that is directed from the n-type side to the p-type side, it follows that there is a difference in potential or voltage across the junction. This voltage difference is called the *built-in potential* or the *contact potential*. It can be calculated from either Eq. (1.31) or Eq. (1.32). Let us use Eq. (1.32). In this equation, p_1 is the hole concentration in the p-type side and p_2 is the hole concentration in the n-type side. By Eq. (1.16) the hole concentration in the p-type side is $p_1 \simeq N_A$, where N_A is the acceptor concentration per m^3. By Eq. (1.15) the hole concentration in the n-type side is $p_2 \simeq n_i^2/N_D$, where n_i is the intrinsic concentration per m^3 and N_D is the donor concentration per m^3. It follows from Eq.

(1.32) that the built-in potential V_B is is given by

$$V_B = V_T \ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{1.33}$$

The same result is obtained from Eq. (1.31).

Example 6 *An open-circuited p-n junction is fabricated from silicon. The acceptor and donor concentrations are $N_A = N_D = 5 \times 10^{20}$ per m^3. The intrinsic concentration is $n_i = 1.5 \times 10^{16}$ per m^3. Solve for the built-in potential at room temperature.*

Solution. The thermal voltage at room temperature is $V_T = 0.0259\,\mathrm{V}$. By Eq. (1.33), the built-in potential has the value

$$V_B = V_T \ln\left(\frac{N_A N_D}{n_i^2}\right) = 0.0259 \ln\left[\frac{\left(5 \times 10^{20}\right)^2}{\left(1.5 \times 10^{16}\right)^2}\right] = 0.539\,\mathrm{V}$$

1.7 The Short-Circuited p-n Junction

Figure 1.14 shows a p-n junction which has metal contacts attached to each end and a short circuiting wire connected between the contacts. From the discussion of the open-circuited p-n junction, it might seem that the built-in potential across the depletion region would cause a current to flow in the external wire. If this happened, the second law of thermodynamics would be violated. This law states that a system in equilibrium with its environment cannot deliver work. To see how this law would be violated, let us assume that a current flows in the short circuiting wire. The current must flow through the semiconductor material which exhibits a finite conductivity σ given by Eq. (1.23). Because heat is generated when a current flows through a conductor of finite conductivity, the system generates energy in the form of heat with no input energy from external sources. This is a clear violation of the second law of thermodynamics.

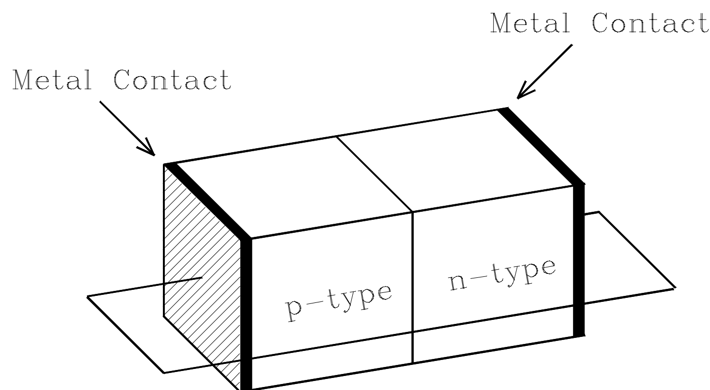

Figure 1.14: Short-circuited p-n junction.

The reason that a current does not flow in the short circuit is because the path through the p-n junction and the wire contains two metal-semiconductor junctions as well as the p-n junction. Like

the p-n junction, both metal-semiconductor junctions exhibit a built-in potential. The algebraic sum of the three built-in potentials is zero so that there is no net voltage to cause a current to flow in the wire and the second law of thermodynamics is not violated.

The characteristics of metal-semiconductor junctions differ primarily from those of p-n junctions in two ways. First, they conduct current well in both directions. Second, the built-in potential does not change when a current flows through the junction. Junctions which have these properties are called *ohmic junctions* or *nonrectifying junctions*. We will see in the following that the p-n junction does not have the properties of an ohmic junction. (Not all metal-semiconductor junctions are ohmic. The Schottky-barrier diode is an example which is described in the following chapter.)

1.8 The Biased p-n Junction

1.8.1 Reverse Bias

Figure 1.15 shows a p-n junction with a dc source connected to it. The polarity of the source is chosen so that the positive terminal is connected to the n-type side and the negative terminal is connected to the p-type side. The current that flows across the junction consists of two components, a diffusion current caused by unequal carrier concentrations on the two sides of the junction and a conduction current caused by the electric field across the junction. With $V_S = 0$, these two currents exactly cancel each other so that the net current is zero.

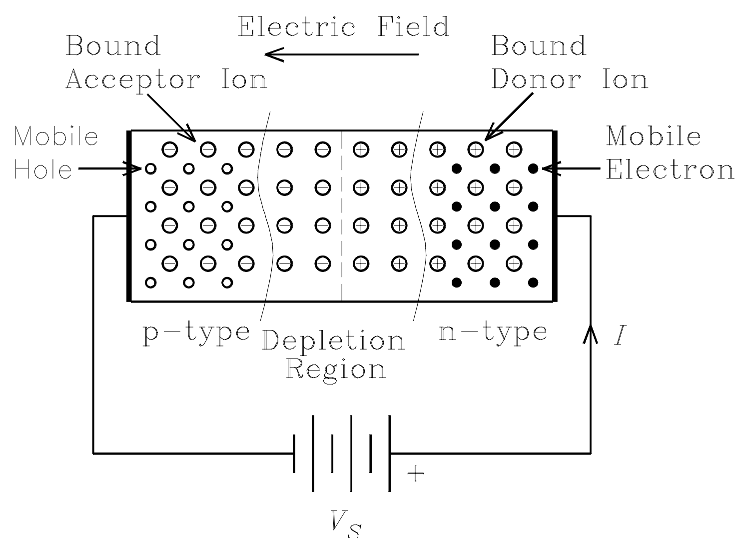

Figure 1.15: Reverse biased p-n junction.

Now let us examine what happens when $V_S > 0$. Because negative charge is attracted by a positive voltage and positive charge is attracted by a negative voltage, both the free electrons in the n-type side and the holes in the p-type side are pulled away from the junction. This causes the width of the depletion region to increase so that there are more uncovered charges on each side of the junction. This is illustrated in Fig. 1.15 compared to Fig. 1.12. The potential across the junction which opposes diffusion is increased by the applied bias to the value $V_B + V_S$, where V_B

is the built-in potential. This is greater than V_B so that the electric field across the junction is increased. Because the diffusion force on the charge carriers is opposed by the force exerted by this electric field, it follows that the diffusion current is decreased by the applied voltage. The diffusion current approaches zero as V_S is increased.

Although the diffusion current goes to zero, the conduction current due to thermally produced hole-electron pairs in the depletion region continues to flow across the junction. When such a hole-electron pair is generated, the electric field across the junction causes the electron to be swept to the n-type side and the hole to be swept to the p-type side. This causes a very small current to flow in the external circuit in the direction indicated in Fig. 1.15. Because the current is so small, the junction is said to be *reverse biased*.

1.8.2 Forward Bias

Figure 1.16 shows the p-n junction with the polarity of the dc source chosen so that the positive terminal is connected to the p-type side and the negative terminal is connected to the n-type side. Because positive charge is repelled by a positive voltage and negative charge is repelled by a negative voltage, both the free electrons in the n-type side and the holes in the p-type side are forced toward the junction. This causes the width of the depletion region to decrease so that there are fewer uncovered charges on each side of the junction. This is illustrated in Fig. 1.16 compared to Fig. 1.12. The potential across the junction which opposes diffusion is decreased by the applied bias to the value $V_B - V_S$, where V_B is the built-in potential. This is less than V_B so that the electric field across the junction is decreased. This decreases the force which opposes diffusion so that the diffusion current increases rapidly as V_S is increased. Because a large current flows, the junction is said to be *forward biased*.

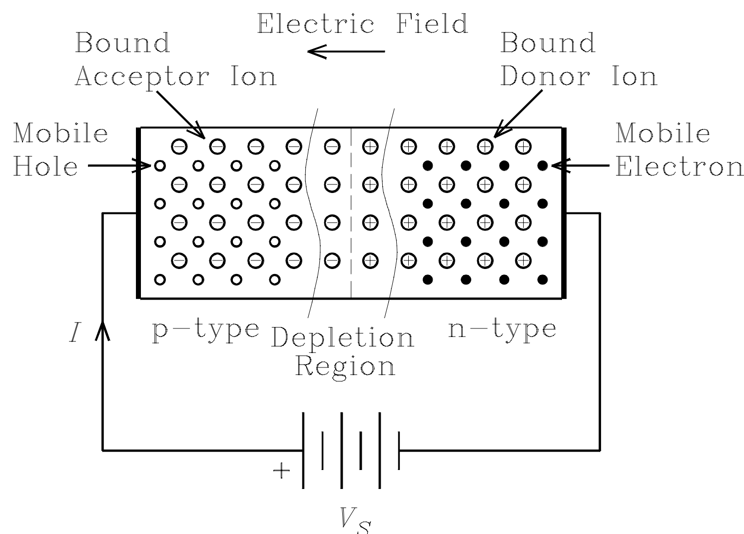

Figure 1.16: Forward biased p-n junction.

It might seem that the potential across the junction which opposes diffusion could be made to go to zero by increasing V_S. Should this happen, the width of the depletion region would go

to zero and the current would become arbitrarily large. This cannot happen in a physical p-n junction because the resistance of the semiconductor material and the resistances of the external metal contacts limit the current. If we denote the sum of the resistances by R, the net potential across the junction which opposes diffusion is $V_B - (V_S - IR) = V_B - V_S + IR$. This relation shows that the voltage drop across R counteracts the effect of V_S on the voltage which opposes diffusion and this in turn limits the current.