



Equilibrium Charge Carrier Statistics in Semiconductors

7

7.1 Introduction

In Chap. 4, we discussed the quantum mechanical states of electrons in a periodic crystal potential and the resulting formation of energy bands. We also introduced the concept of effective mass, that of holes, and the Fermi energy which provides an easy way to differentiate a semiconductor from a metal.

In semiconductor devices, most of the properties of interest have their origins in the electrons in the conduction band and the holes in the valence band. Two major functions are important in understanding the behavior of these electrons and holes: the density of states and the Fermi-Dirac distribution function, both of which have been discussed in Chaps. 4 and 5. In this chapter, we will establish the basic relations and formalism for the distribution of electrons in the conduction band and holes in the valence band at thermal equilibrium. We will also introduce the notion of doping and extrinsic semiconductors, in contrast to pure or intrinsic semiconductors.

7.2 Density of States

In Chap. 5, we calculated the density of states of electrons of the conduction band in a three-dimensional semiconductor to be:

$$g_c(E) = \frac{V}{2\pi^2} \left(\frac{2m_e}{\hbar^2} \right)^{3/2} (E - E_C)^{1/2} \quad (7.1)$$

where m_e is the electron effective mass in the conduction band, E_C is the bottom of the conduction band, and V is the volume of the crystal considered. The subscript “c” in g_c indicates that we are considering the conduction band. This expression was calculated for a single band minimum and is valid for direct-gap semiconductors, such as GaAs, where the conduction band minimum occurs at the zone center.

However, in the case of many other semiconductors, one has to take into account the degeneracy or number g_d of equivalent conduction band minima in the first Brillouin zone.

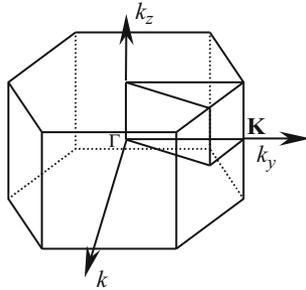
For example, we saw in Fig. 5.17a that the conduction band minimum in Ge occurred along the $\langle 111 \rangle$ direction. As there are eight equivalent $\langle 111 \rangle$ directions, there are eight equivalent conduction band minima in Ge. However, because the minima occur exactly at the boundary of the first Brillouin zone, each minimum is shared with two neighboring zones and therefore only contributes one half to the density of states. Thus $g_{\text{deg}} = 4$, i.e., the expression in Eq. (5.52) needs to be multiplied by a factor 4. In addition, we also saw in Fig. 5.17b that the conduction band minimum in Si occurs at $k \approx 0.8(2\pi/a)$ in the first Brillouin zone along the $\langle 100 \rangle$ direction. Since the $\langle 100 \rangle$ direction has a sixfold symmetry, this gives rise to six equivalent conduction band minima within the first Brillouin zone, and $g_d = 6$ because the minimum is strictly inside the first Brillouin zone. The expression in Eq. (7.1) then needs to be multiplied by 6. Finally, for GaAs, as shown in Fig. 5.17c, the conduction band minimum occurs at the zone center, and the expression in Eq. (7.1) remains unchanged, i.e., $g_d = 1$.

In other words, the full density of states of electrons in the conduction band is ($E > E_C$):

$$g_c(E) = \frac{V}{2\pi^2} g_d \left(\frac{2m_e}{\hbar^2} \right)^{3/2} (E - E_C)^{1/2} \quad (7.2)$$

Example

- Q GaN has the wurtzite crystal structure. The first Brillouin zone is shown in the figure below. From the calculation of the band structure of GaN, it can be seen that there is a shallow conduction band minimum at the symmetry point K in the reciprocal lattice. To calculate the density of states given by the expression $g_c(E) = \frac{V}{2\pi^2} g_d \left(\frac{2m_e}{\hbar^2} \right)^{3/2} (E - E_C)^{1/2}$, what is the degeneracy factor g_d which should be used?



A The point K is equally shared by three adjacent Brillouin zones. Because the first Brillouin zone has sixfold symmetry, there are six equivalent points K in the zone. This leads to a total degeneracy of : $6 \times \frac{1}{3} = 2$.

The value of the electron effective mass m_e was determined in Eq. (5.27), in the simple case of a one-dimensional crystal, as the curvature of the conduction band or, in other words, the second derivative of the energy spectrum $E(k)$ such that $E(k)$ can be approximated as:

$$E(k) \approx \frac{\hbar^2}{2m_e} k^2 \quad (7.3)$$

In the more general case of a three-dimensional crystal, the effective mass is a 3×3 matrix, and each element is a function of the direction in which the two derivatives of the energy spectrum $E(\vec{k})$ are performed, k_x , k_y , or k_z .

If the energy spectrum can be approximated as:

$$E(\vec{k}) \approx \frac{\hbar^2}{2} \left(\frac{k_x^2}{m_{xx}} + \frac{k_y^2}{m_{yy}} + \frac{k_z^2}{m_{zz}} \right) \quad (7.4)$$

where m_{xx} , m_{yy} , and m_{zz} correspond to the values of the second partial derivatives in the k_x , k_y , and k_z directions, respectively; then the electron effective mass m_e that is considered in Eq. (7.3) is the average of these three masses and is given by:

$$m_e = (m_{xx}m_{yy}m_{zz})^{1/3} \quad (7.5)$$

In the particular case when the energy spectrum can be approximated as:

$$E(\vec{k}) \approx \frac{\hbar^2}{2} \left(\frac{(k_x^2 + k_y^2)}{m_t} + \frac{k_z^2}{m_l} \right) \quad (7.6)$$

where m_t and m_l are customarily called the transverse electron effective mass and the longitudinal electron effective mass, respectively; then the electron effective mass m_e that is considered in Eq. (7.5) is the average of these three masses and is given by:

$$m_e = (m_t^2 m_l)^{1/3} \quad (7.7)$$

A similar relation can be obtained for the electronic density of states in the valence band ($E_V < E$):

$$g_v(E) = \frac{V}{2\pi^2} \left(\frac{2m_h}{\hbar^2} \right)^{3/2} (E_V - E)^{1/2} \quad (7.8)$$

where m_h is the hole effective mass which accounts for the curvature of the valence band and E_V is the top of the valence band. In this expression, there is no degeneracy factor from crystal symmetry because the top of the valence band is unique and always occurs at the center of the first Brillouin zone.

We saw in Sect. 5.4 that the valence band of a semiconductor is composed of two main subbands, the heavy-hole and light-hole bands, each with a different curvature and thus with their own hole effective masses: m_{hh} and m_{lh} , for the heavy-hole effective mass and light-hole effective mass, respectively. As a result, the hole effective mass m_h that is considered in Eq. (7.7) is the following average of these two masses:

$$m_h = \left(m_{hh}^{3/2} + m_{lh}^{3/2} \right)^{2/3} \quad (7.9)$$

7.3 Effective Density of States (Conduction Band)

As discussed in Sub-sect. 5.2.8, the density of states merely provides information about the allowed energy states. To obtain the concentration of electrons in the conduction band, we must multiply this density of states with the Fermi-Dirac distribution (Eq. (5.28)) which gives the probability of occupation of an energy state:

$$n = \frac{1}{V} \int_{E_C}^{\infty} g_c(E) f_e(E) dE \quad (7.10)$$

Expanding this expression using Eq. (5.52) and Eq. (5.28), we get:

$$n = \frac{g_d}{2\pi^2} \left(\frac{2m_e}{\hbar^2} \right)^{3/2} \int_{E_C}^{\infty} \frac{(E - E_C)^{1/2}}{\exp\left(\frac{E - E_F}{k_b T}\right) + 1} dE \quad (7.11)$$

Making the change of variable $y = \frac{E - E_C}{k_b T}$, and thus $dy = \frac{1}{k_b T} dE$, the previous integral becomes:

$$\int_{E_C}^{\infty} \frac{(E - E_C)^{1/2}}{\exp\left(\frac{E - E_F}{k_b T}\right) + 1} dE = (k_b T)^{3/2} \int_0^{\infty} \frac{y^{1/2}}{\exp\left(y - \frac{E_F - E_C}{k_b T}\right) + 1} dy \quad (7.12)$$

We can define the Fermi-Dirac integral as in Eq. (5.56):

$$F_{\frac{1}{2}}(x) = \frac{2}{\sqrt{\pi}} \int_0^{\infty} \frac{y^{1/2}}{1 + \exp(y - x)} dy \quad (7.13)$$

using:

$$x = \frac{E_F - E_C}{k_b T} \quad (7.14)$$

Equation (7.12) can be rewritten as:

$$\int_{E_C}^{\infty} \frac{(E - E_C)^{1/2}}{\exp\left(\frac{E - E_F}{k_b T}\right) + 1} dE = (k_b T)^{3/2} \frac{\sqrt{\pi}}{2} F_{1/2}\left(\frac{E_F - E_C}{k_b T}\right) \quad (7.15)$$

and therefore Eq. (7.11) becomes:

$$n = \frac{g_d}{2\pi^2} \left(\frac{2m_e}{\hbar^2}\right)^{3/2} (k_b T)^{3/2} \frac{\sqrt{\pi}}{2} F_{1/2}\left(\frac{E_F - E_C}{k_b T}\right) \quad (7.16)$$

Remembering that $\hbar = \frac{h}{2\pi}$, this can be simplified as:

$$n = 2g_d \left(\frac{2\pi k_b T m_e}{h^2}\right)^{3/2} F_{1/2}\left(\frac{E_F - E_C}{k_b T}\right) \quad (7.17)$$

or:

$$n = N_c F_{1/2}\left(\frac{E_F - E_C}{k_b T}\right) \quad (7.18)$$

with:

$$N_c = 2g_d \left(\frac{2\pi k_b T m_e}{h^2}\right)^{3/2} \quad (7.19)$$

N_c is called the effective conduction band density of states. The Fermi-Dirac integral defined in Eq. (7.13) is often approximated with simpler expressions. One commonly encountered situation is when $E_C - E_F \gg k_b T$. A semiconductor in this situation is called a non-degenerate semiconductor. Let us give a numerical example. At room temperature ($T = 300$ K), we have $k_b T = 25.9$ meV. Therefore, we can consider that we are in the presence of a non-degenerate semiconductor when the Fermi energy E_F is away from the bottom of the conduction band E_C by a few times 25.9 meV. This is illustrated in Fig. 7.1a. For most of the practical calculations, a distance of $3k_b T$ or more, i.e., $E_C - E_F \geq 3k_b T$, is sufficient.

This approximation means that the Fermi energy is rather far from the bottom of the conduction band and inside the bandgap and that $x \ll -1$ in Eq. (7.13). Therefore, the exponential function dominates in the denominator for all positive values of $y > 0$, i.e., $1 + \exp(y - x) \approx \exp(y - x)$. Thus:

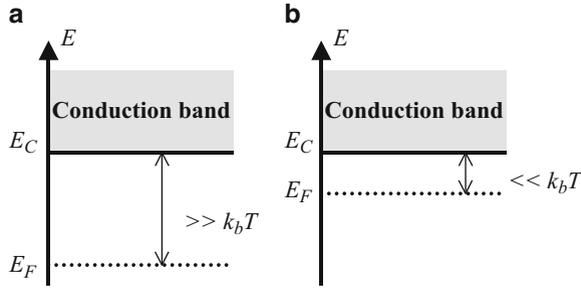


Fig. 7.1 Illustration of the position of the Fermi level with respect to the conduction band (a) in a non-degenerate n-type semiconductor: The Fermi energy is far from the edge of the conduction band. (b) In a degenerate semiconductor n-type semiconductor, the Fermi energy is close to the edge of the conduction band

$$F_{\frac{1}{2}}(x) \approx \frac{2}{\sqrt{\pi}} \int_0^{\infty} \frac{y^{1/2} dy}{\exp(y-x)} = \frac{2}{\sqrt{\pi}} e^x \int_0^{\infty} y^{1/2} e^{-y} dy \quad (7.20)$$

The integral on the right hand side can be transformed by integrating by parts:

$$\begin{aligned} \int_0^{\infty} y^{1/2} e^{-y} dy &= [-y^{1/2} e^{-y}]_0^{\infty} + \frac{1}{2} \int_0^{\infty} y^{-1/2} e^{-y} dy \\ &= \frac{1}{2} \int_0^{\infty} y^{-1/2} e^{-y} dy \end{aligned}$$

Making now the change of variable $Y = y^{1/2}$, and thus $dY = \frac{1}{2} y^{-1/2} dy$, we get the well-known integral:

$$\frac{1}{2} \int_0^{\infty} y^{-1/2} e^{-y} dy = \frac{1}{2} \int_0^{\infty} e^{-Y^2} dY = \frac{\sqrt{\pi}}{2}$$

Substituting in Eq. (7.20), we obtain for a non-degenerate semiconductor:

$$F_{\frac{1}{2}}(x) \approx e^x$$

and from Eqs. (7.18) and (7.20):

$$n \approx N_c \exp\left(\frac{E_F - E_C}{k_b T}\right) \quad (7.21)$$

This expression is much simpler than Eq. (7.16) and is more amenable for calculations. However, when the Fermi energy is close to or even higher than the

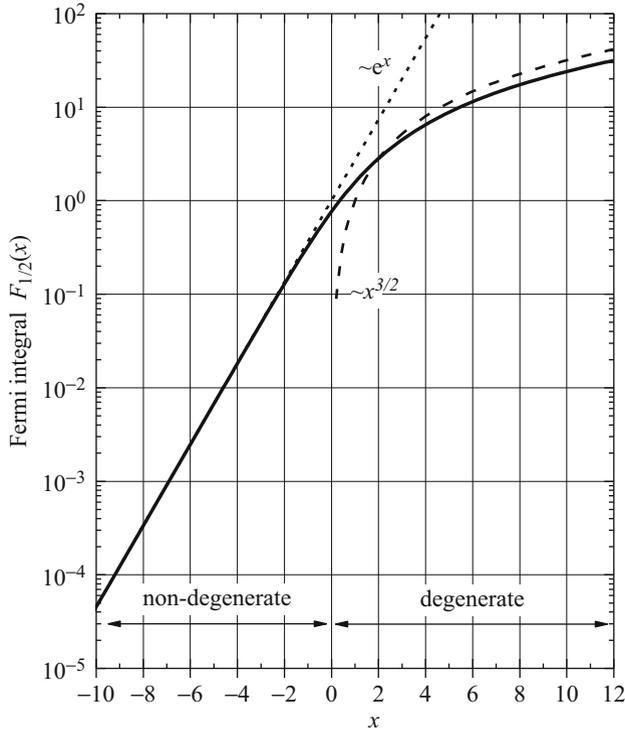


Fig. 7.2 The Fermi integral of order one half and its approximations

bottom of the conduction band, we have a so-called degenerate semiconductor, we cannot make this approximation anymore, and the Fermi-Dirac integral has to be used.

An extreme case is when $E_F - E_C \gg k_b T$, corresponding to *highly degenerate* semiconductors, in which the Fermi level lies deeply inside the conduction band. Electrical properties of such semiconductors are similar to those of metals. At this condition, the Fermi-Dirac integral can be approximated as:

$$F_{\frac{1}{2}}(x) \approx x^{3/2}$$

Figure 7.2 shows the plots of the Fermi integral and the two approximations mentioned above. The exponential approximation or the 3/2 power approximation agrees very well with the Fermi integral when $x \ll -1$ or $x \gg 1$. However, when $x \sim 0$, the Fermi-Dirac integral has to be used.

Fortunately, we are almost exclusively concerned with non-degenerate semiconductors. For example, InSb has a bandgap of 0.17 eV at 300 K, which is one of the smallest bandgaps among all the semiconductors. Assume InSb is pure and perfect (or so-called intrinsic, see Sect. 7.6), the Fermi energy is approximately

in the middle of the bandgap, $E_C - E_F \approx E_g/2$, which is about 85 meV at 300 K. Note that $k_bT = 25.9$ meV; the condition $E_C - E_F \geq 3k_bT$ is satisfied. Thus the exponential form can be used. Most of the semiconductors have a larger bandgap, which means the $3k_bT$ condition is valid at room temperature.

7.4 Effective Density of States (Valence Band)

A similar derivation can be performed for the concentration or density of holes p in the valence band:

$$p = \frac{1}{V} \int_{-\infty}^{E_V} g_v(E) f_h(E) dE \quad (7.22)$$

which we obtained from Eq. (7.10) after replacing the density of states with that in the valence band and the limit of integration for an energy below the top of the valence band E_V . Moreover, the Fermi-Dirac distribution $f_e(E)$ has been replaced with (see Eq. (5.58)):

$$f_h(E) = [1 - f_e(E)] = \frac{1}{\exp\left(\frac{E_F - E}{k_bT}\right) + 1} \quad (7.23)$$

which gives the probability of the state at energy E not to be occupied by an electron and thus to be occupied by a hole.

Expanding Eq. (7.22) using Eqs. (7.8) and (7.23), we get:

$$p = \frac{1}{2\pi^2} \left(\frac{2m_h}{\hbar^2}\right)^{3/2} \int_{-\infty}^{E_V} \frac{(E_V - E)^{1/2}}{\exp\left(\frac{E_F - E}{k_bT}\right) + 1} dE \quad (7.24)$$

Using the change of variable $y = \frac{E_V - E}{k_bT}$, thus $dy = -\frac{1}{k_bT} dE$, and:

$$x = \frac{E_V - E_F}{k_bT} \quad (7.25)$$

in the previous integral and identifying it with the Fermi-Dirac integral, we obtain a relation similar to Eq. (7.17) for p :

$$p = 2 \left(\frac{2\pi k_b T m_h}{\hbar^2}\right)^{3/2} F_{\frac{1}{2}}\left(\frac{E_V - E_F}{k_b T}\right) \quad (7.26)$$

or:

$$p = N_v F_{\frac{1}{2}}\left(\frac{E_V - E_F}{k_b T}\right) \quad (7.27)$$

where:

$$N_v = 2 \left(\frac{2\pi k_b T m_h}{h^2} \right)^{3/2} \quad (7.28)$$

is called the effective valence band density of states.

Example

Q Find the ratio of the heavy-hole concentration to the light-hole concentration for GaAs.

A We know that the hole concentration is related to the hole effective mass through:

$$p = 2 \left(\frac{2\pi k_b T m_h}{h^2} \right)^{3/2} F_{1/2} \left(\frac{E_V - E_F}{k_b T} \right).$$

The Fermi-Dirac integral is the same for the heavy-hole and light-hole bands, and the only difference comes from the effective masses. Therefore, we can write:

$$\frac{p_{hh}}{p_{lh}} = \left(\frac{m_{hh}}{m_{lh}} \right)^{3/2}. \text{ In GaAs, this ratio is: } \frac{p_{hh}}{p_{lh}} = \left(\frac{0.45}{0.082} \right)^{3/2} = 12.86$$

Similar to what we saw in Sect. 7.3, the general expression in Eq. (7.24) can be simplified in the case of a non-degenerate semiconductor for which $E_F - E_V \gg k_b T$. This situation is of most interest and is illustrated in Fig. 7.3a. It corresponds to the one where the Fermi energy is rather far from the valence band and inside the bandgap.

In this situation, the concentration of holes has a simplified expression similar to Eq. (7.18):

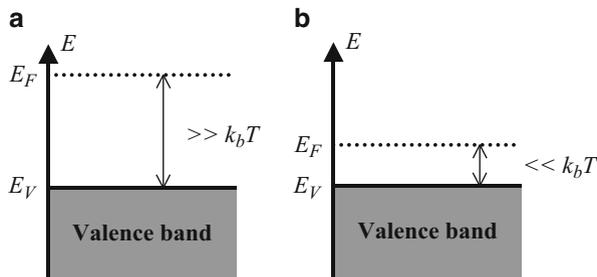


Fig. 7.3 Illustration of the position of the Fermi level with respect to the valence band (a) in a non-degenerate p -type semiconductor: The Fermi energy is far from the edge of the valence band. (b) In a degenerate p -type semiconductor, the Fermi energy is close to the edge of the valence band

$$p \approx N_v \exp\left(\frac{E_V - E_F}{k_b T}\right) \quad (7.29)$$

7.5 Mass Action Law

We saw that a non-degenerate semiconductor has its Fermi energy far away from both the bottom of the conduction band and the top of the valence band, by about a few times $k_b T$ (25.9 meV at room temperature). This situation is more often encountered in practice than one may believe, and most of the discussion from now will therefore be in this approximation unless stated otherwise.

An important parameter is the product of n and p given in Eqs. (7.21) and (7.29) by:

$$\begin{aligned} np &= N_c \exp\left(\frac{E_F - E_c}{k_b T}\right) N_v \exp\left(\frac{E_V - E_F}{k_b T}\right) \\ &= N_c N_v \exp\left(\frac{E_V - E_c}{k_b T}\right) \end{aligned}$$

or:

$$np = N_c N_v \exp\left(-\frac{E_g}{k_b T}\right) \quad (7.30)$$

where $E_g = E_c - E_v$ is the bandgap energy of the semiconductor. This relation is very important, as it is valid for *any* value of n or p . This relation is usually called the mass action law. However, it does not hold in the degenerate semiconductor case. It is common practice to introduce the intrinsic carrier concentration, n_i , which is defined as:

$$n_i^2 = np = N_c N_v \exp\left(-\frac{E_g}{k_b T}\right) \quad (7.31)$$

This parameter is a function of the semiconductor effective masses and the temperature. This concentration is qualified as “intrinsic” because for an intrinsic semiconductor, the number of electrons and holes are equal, i.e., $n = p$, and we thus have from the previous relation:

$$n = p = n_i = \sqrt{N_c N_v} \exp\left(-\frac{E_g}{2k_b T}\right) \quad (7.32)$$

Example

Q Calculate the intrinsic electron concentration for undoped GaAs at room temperature (300 K).

- A For a homogeneous non-degenerate semiconductor, like undoped GaAs, the mass action law gives the intrinsic electron concentration as:

$$\begin{aligned} n_i &= \sqrt{4g_d \left(\frac{2\pi k_b T m_e}{h^2} \right)^{3/2} \left(\frac{2\pi k_b T m_h}{h^2} \right)^{3/2}} \exp\left(-\frac{E_g}{2k_b T}\right) \\ &= 2 \left(\frac{2\pi k_b T m_0}{h^2} \right)^{3/2} \sqrt{g_d \left(\frac{m_e m_h}{m_0 m_0} \right)^{3/2}} \exp\left(-\frac{E_g}{2k_b T}\right) \end{aligned}$$

where E_g is the bandgap of GaAs (1.424 eV). For GaAs, the degeneracy factor g_d is equal to 1 because the conduction band minimum is at the center of the Brillouin zone. In addition, the hole effective mass m_h is calculated from the heavy-hole and light-hole effective masses: $m_h^{3/2} = m_{hh}^{3/2} + m_{hl}^{3/2}$. We therefore get:

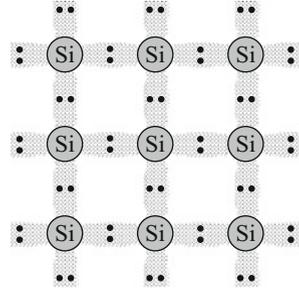
$$\begin{aligned} n_i &= 2 \left(\frac{2\pi \times (1.38066 \times 10^{-23}) \times 300 \times (0.91095) \times 10^{-30}}{(6.62617 \times 10^{-34})^2} \right)^{3/2} \\ &\times \sqrt{(0.067)^{3/2} (0.45^{3/2} + 0.082^{3/2})} \\ &\times \exp\left(-\frac{1.424 \times 1.60218 \times 10^{-19}}{2 \times (1.38066 \times 10^{-23}) \times 300}\right) \\ &= 2.06 \times 10^{12} \text{m}^{-3} \\ &= 2.06 \times 10^6 \text{cm}^{-3} \end{aligned}$$

7.6 Doping: Intrinsic Versus Extrinsic Semiconductor

The energy band structures of the semiconductors that have been discussed so far corresponded to those of an intrinsic semiconductor, which is a pure and perfect semiconductor crystal. At a temperature equal to the absolute zero (0 K), the valence band of such a crystal is completely filled with electrons, and there is no electron in the conduction band. Indeed, we saw that the Fermi energy of a semiconductor lies within a forbidden energy gap (Sub-sect. 5.2.7). Since the Fermi-Dirac distribution function has an exact step shape at $T = 0$ K (Fig. 5.12), there is no electron with an energy $E > E_F$, including the conduction band, and all the electrons are located at an energy $E < E_F$.

This phenomenon directly results from the fact that the outer shell of each constituent atom of a semiconductor is fully filled with four electrons. Counting the number in the four shared bonds then gives a total of eight electrons. For example, in the case of a silicon crystal, illustrated in Fig. 7.4, each Si atom is bonded to four neighboring Si atoms. A Si atom originally has four electrons in its outer shell (it is in the column IV of the periodic table), each of which is shared with

Fig. 7.4 Schematic of a Si semiconductor crystal showing the distribution of electrons in the outer shell of each Si atom. Each Si atom has eight electrons in this shell: four from its own outer shell and one from each of the four nearest Si atoms to which it is covalently bonded to



one different neighboring atom. Every Si atom has therefore a total of eight electrons: Its original four electrons and one electron from each of the four neighboring Si atoms.

We thus see that all the outer shell electrons are shared into bonds, and thus there is no extra free electron which can move. Moreover, all the outer shell “spots” are filled with electrons; therefore there is no room for an electron to move to if displaced by a field. As a result, the electrical conductivity of a pure semiconductor is “low” (only excited states can conduct). This is why a pure semiconductor is an insulator at the absolute zero temperature.

In order to either increase the number of free electrons or increase the number of “spots” (empty energy levels) where a potential electron can move into, we need to replace some of the Si atoms with other elements, called dopants, which are not isoelectronic to it, i.e., not with the same number of outer shell electrons. This process is called doping, which results in an extrinsic semiconductor. A dopant is thus an impurity added to the semiconductor crystal. Because the dopant replaces or substitutes a Si atom, it is called a substitutional dopant. The concentration of such dopants typically introduced in a semiconductor is in the range of 10^{15} – 10^{19} cm^{-3} , which is low in comparison with the concentration of atoms in a crystal (typically $\sim 10^{22}$ cm^{-3}). There are two types of doping, *n*-type doping and *p*-type doping, depending on the nature of the dopant introduced. Such a dopant can be introduced intentionally or unintentionally during the synthesis of the semiconductor crystal.

The *n*-type doping is achieved by replacing a Si atom with an atom with *more* electrons in the outer shell. This can be achieved, for example, by using phosphorus (P), an element from the column V of the periodic table, which has five electrons in its outer shell. The result is shown in Fig. 7.5.

As we can see, four of the electrons in the outer shell of the P atom are involved in covalent bonds with its four neighboring Si atoms. The fifth electron is therefore free to move in space. The P atom is therefore called a donor in silicon because it can give away an electron which can in turn participate in electrical conductivity phenomena. Once an electron is given away, the phosphor atom becomes a positively charged ion and is then called an ionized donor. This ionization process is generally achieved through thermal excitation of an electron from the outer shell of the donor atom.

Because the dopant creates a perturbation to the periodicity of the crystal lattice, it gives rise to additional energy levels in the bandgap. When the dopant concentration

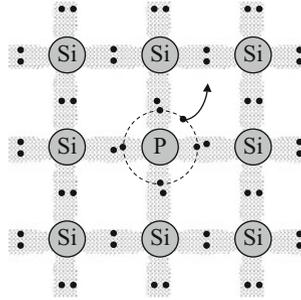
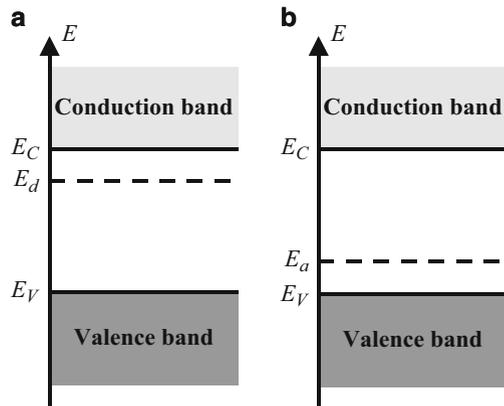


Fig. 7.5 Schematic of a Si semiconductor crystal with one Si atom replaced by a P atom to achieve n -type doping. The dotted circle symbolizes the outer shell of the P atom which contains five electrons. Because the fifth electron does not contribute to the bonding, it can be free (ionized) to move inside the crystal. P is thus called a donor

Fig. 7.6 Schematic of the energy levels introduced by (a) a donor or (b) an acceptor dopant in a semiconductor crystal. The energy level of a donor is closer to the edge of the conduction band, whereas that of an acceptor is closer to the edge of the valence band



is low in comparison with the density of crystal atoms, the dopant energy level can be considered as isolated, i.e., there is no energy band associated with it. We can then talk about a donor energy level E_d , as shown in Fig. 7.6a. Moreover, because the extra electron around the P atom is easily ionized, it has a small binding energy, with respect to the conduction band. The energy of the donor electron E_d is closer to the conduction band than the valence band. The ionization energy of the dopant is the difference $E_C - E_d$.

The other type of doping, p -type doping, is achieved by replacing a Si atom with an atom with *fewer* electrons in the outer shell. This can be achieved, for example, by using gallium (Ga), an element from the column III of the periodic table, which has three electrons in its outer shell. The result is shown in Fig. 7.7.

As we can see, all three electrons in the outer shell of the Ga atom are involved in covalent bonds with three of its four neighboring Si atoms. There thus remains an open location that can be filled with an electron. The Ga atom is therefore called an acceptor in silicon because it can “accept” or “capture” an extra electron from a

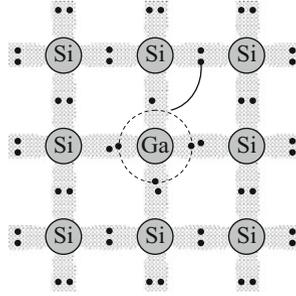
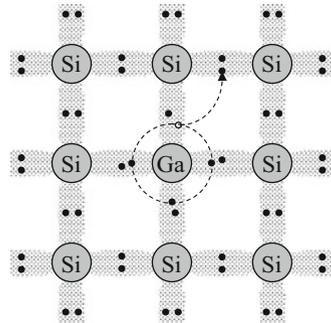


Fig. 7.7 Schematic of a Si semiconductor crystal with one Si atom replaced by a Ga atom to achieve p -type doping. The dotted circle symbolizes the outer shell of the Ga atom which contains three electrons. The Ga atom can accept one more electron from a neighboring bond. Ga is thus called an acceptor

Fig. 7.8 Schematic showing the movement of a hole in a Si semiconductor crystal doped p -type using a Ga atom. The hole is represented by an open circle. When the Ga atom accepts an electron, the process can be equivalently viewed as the Ga atom releasing a hole inside the crystal



neighboring covalent bond, thus leaving a new available location for electron capture. Once an electron is captured, the gallium atom becomes a negatively charged ion and is then called an ionized acceptor. This movement of electrons is involved in electrical conductivity phenomena. Remembering the concept of holes discussed in Sub-sect. 5.3.3, we can see that this electron movement is equivalent to the movement of a hole in the opposite direction, as illustrated in Fig. 7.8. The Ga atom, an acceptor (of electrons) in silicon, can then be also considered as a donor of holes.

Here again, the p -type dopant is a perturbation of the periodicity of the crystal lattice and leads to additional localized energy levels (i.e., not bands) in the bandgap at E_a , which is called acceptor energy level, as shown Fig. 7.6b. Because the Ga atom easily captures an electron, E_a is closer to the valence band than the conduction band. The ionization energy of the p -type dopant is the difference $E_a - E_V$.

A semiconductor may contain donors (with a concentration N_D) and acceptors (with a concentration N_A) at the same time. We then talk about compensation and say that the semiconductor is compensated. The overall behavior of this semiconductor depends on the relative difference between N_D and N_A . In either case of n -type

Table 7.1 Dopants and ionization energies for (a) Si, (b) Ge, (c) GaAs (Sze 1981; Wolfe et al. 1989), and (d) InP (<http://www.ioffe.ru/SVA/NSM/Semicond/InP/index.html>)

(a) Si		
Impurity	Type	Ionization energy (meV)
P	Donor	45.31
As	Donor	53.51
Sb	Donor	42.51
B	Acceptor	45
Al	Acceptor	57
Ga	Acceptor	65
(b) Ge		
Impurity	Type	Ionization energy (meV)
P	Donor	12.76
As	Donor	14.04
Sb	Donor	10.19
B	Acceptor	10.47
(c) GaAs		
Impurity	Type	Ionization energy (meV)
Si	Donor	5.854
Ge	Donor	5.908
S	Donor	5.89
Be	Acceptor	30
Mg	Acceptor	30
Zn	Acceptor	31.4
C	Acceptor	26.7
(d) InP		
Impurity	Type	Ionization energy (meV)
Si	Donor	5.7
S	Donor	5.7
Sn	Donor	5.7
Be	Acceptor	30
Mg	Acceptor	30
Zn	Acceptor	35

and/or *p*-type doping, the mass action law expressed in Eq. (7.30) remains valid as long as we have a non-degenerate semiconductor.

Table 7.1 lists the most common dopants with their ionization energies for the following semiconductors: Si, Ge, GaAs, and InP.

7.7 Charge Neutrality

A semiconductor crystal, be it intrinsic or extrinsic, must be electrically neutral at a macroscopic scale. Indeed, even if dopants are introduced, they are electrically neutral, and therefore the semiconductor crystal remains globally neutral too. As the dopants get ionized, they create mobile electrons and holes in the crystal. But,

there is no persistent accumulation of electrical charges. Even in a compensated semiconductor, overall charge neutrality remains.

Before mathematically expressing the electrical neutrality condition, we must first count all the electrical charges present in the crystal. The negative charges include the electrons in the conduction band, with a concentration n , and the ionized acceptors with a concentration N_A^- . The positive charges include the holes in the valence band, with a concentration p , and the ionized donors with a concentration N_D^+ . The charge neutrality relation can then be written as:

$$n + N_A^- = p + N_D^+ \quad (7.33)$$

For a given semiconductor crystal, the concentrations n and p solely depend on the Fermi energy E_F through Eqs. (7.21) and (7.29) in the non-degenerate case or Eqs. (7.18) and (7.27) in the general case. The concentrations of ionized donors N_D^+ and acceptors N_A^- depend also on the Fermi energy for a given dopant nature, the temperature T , and total concentration as follows:

$$\frac{N_D^+}{N_D} = \frac{1}{2\exp\left(\frac{E_F - E_d}{k_b T}\right) + 1} \quad (7.34)$$

$$\frac{N_A^-}{N_A} = \frac{1}{4\exp\left(\frac{E_a - E_F}{k_b T}\right) + 1} \quad (7.35)$$

where E_F is the Fermi energy, E_d and E_a are the donor and acceptor energy levels in the bandgap, respectively, and N_D and N_A are the total donor and acceptor concentrations, respectively. The factor 2 in Eq. (7.34) arises because the donor atom can in practice only be singly occupied by an electron (electron-electron repulsion will prevent double occupation), and the factor 4 in Eq. (7.35) arises for the same reason and the fact that there are two degenerate subbands in the valence band at the center of the Brillouin zone: the heavy-hole band and the light-hole band (Sub-sect. 5.4.3). Similar to the Fermi-Dirac distribution, Eqs. (7.34) and (7.35) are derived from statistical physics.

The charge neutrality equation is a very important property because it gives an implicit equation which can be used to determine the Fermi energy. Once the Fermi energy is determined, the concentration of electrons in the conduction band and that of holes in the valence band can be readily calculated through Eqs. (7.21) and (7.29) in the non-degenerate case or Eqs. (7.18) and (7.27) in the general case.

7.8 Fermi Energy as a Function of Temperature

An example of such calculation is given here, first for an intrinsic and then for an n -type extrinsic and non-degenerate semiconductor.

In the intrinsic case, we assume there is no dopant, i.e., the total concentration of dopant is $N_D = N_A = 0$. Substituting in Eq. (7.33), we therefore obtain Eq. (7.32) again. Now, by identifying n in Eqs. (7.21) and (7.32), we can write an expression for the Fermi energy:

$$n = \sqrt{N_c N_v} \exp\left(-\frac{E_g}{2k_b T}\right) = N_c \exp\left(\frac{E_F - E_C}{k_b T}\right)$$

which becomes, knowing that $E_g = E_C - E_V$:

$$\exp\left(\frac{E_F}{k_b T}\right) = \sqrt{\frac{N_v}{N_c}} \exp\left(\frac{E_C + E_V}{2k_b T}\right) \quad (7.36)$$

After taking the logarithm of this relation:

$$\frac{E_F}{k_b T} = \frac{E_C + E_V}{2k_b T} + \ln\left(\sqrt{\frac{N_v}{N_c}}\right)$$

or:

$$E_F = \frac{E_C + E_V}{2} + \frac{1}{2} k_b T \ln\left(\frac{N_v}{N_c}\right) \quad (7.37)$$

This equation shows that the Fermi energy in an intrinsic semiconductor lies near the middle of the bandgap and is offset by an amount that varies with temperature. At the absolute zero temperature, the Fermi energy is exactly at the middle of the bandgap.

Example

Q Determine how far the Fermi energy is from the middle of the bandgap of GaAs at 296 K.

A The Fermi energy is given the expression: $E_F = \frac{E_C + E_V}{2} + \frac{1}{2} k_b T \ln\left(\frac{N_v}{N_c}\right)$. The energy difference between the Fermi energy and the middle of the bandgap is therefore given by the logarithm function, $E_F - \frac{E_C + E_V}{2} = \frac{1}{2} k_b T \ln\left(\frac{N_v}{N_c}\right)$, which is given by the ratio: $\frac{1}{2} k_b T \ln\left(\frac{1}{g_d} \left(\frac{m_h}{m_c}\right)^{3/2}\right)$. In GaAs, the degeneracy factor g_d is equal to 1 because the conduction band minimum is at the center of the Brillouin zone. In addition, the hole effective mass m_h is calculated from the heavy-hole and light-hole effective masses: $m_h^{3/2} = m_{hh}^{3/2} + m_{lh}^{3/2}$. This leads to:

$$\begin{aligned}
 E_F - \frac{E_C + E_V}{2} &= \frac{1}{2} k_b T \ln \left(\frac{m_{hh}^{3/2} + m_{lh}^{3/2}}{m_e^{3/2}} \right) = \frac{1}{2} \\
 &\times 1.38066 \times 10^{-23} \times 296 \times \ln \left(0.45^{3/2} + \frac{0.082^{3/2}}{0.067^{3/2}} \right) \\
 &= 6.00 \times 10^{-21} \text{ J} = 37.4 \text{ meV}
 \end{aligned}$$

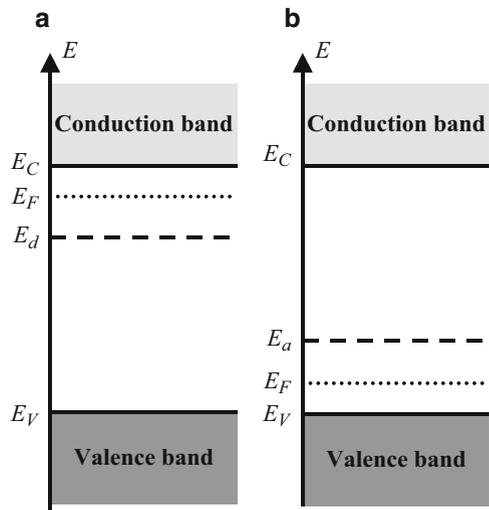
For an extrinsic semiconductor, an expression similar to Eq. (7.37) cannot be easily obtained, because one needs to estimate the concentrations of ionized donors (N_D^+) or acceptors (N_A^-) as a function of the total concentrations, which is beyond the scope of this textbook. Nevertheless, the following discussion will enable us to qualitatively understand the variation of the Fermi energy as a function of temperature.

We know that, at the absolute zero temperature ($T = 0 \text{ K}$), the Fermi energy E_F is such that all the electrons have an energy below E_F and no electron has an energy higher than E_F .

Therefore, in an n -type doped semiconductor at $T = 0 \text{ K}$, the Fermi energy is located between E_C and E_d , as illustrated in Fig. 7.9a, which means that the Fermi energy is much closer to the bottom of the conduction band than in the case of an intrinsic semiconductor. This proximity has the very important consequence that the concentration of electrons in the conduction band is much larger than for an intrinsic semiconductor, as a result of the shape of the Fermi-Dirac distribution shown in Fig. 5.12, when the temperature is raised. These electrons can easily participate in electrical conduction phenomena.

By contrast, in a p -type doped semiconductor at $T = 0 \text{ K}$, the Fermi energy E_F is located between E_V and E_a , as illustrated in Fig. 7.9b, which means that the Fermi energy is much closer to the top of the valence band than in the case of an intrinsic

Fig. 7.9 Position of the Fermi energy at $T = 0 \text{ K}$ in (a) an n -type semiconductor is located between the donor energy level and the bottom of the conduction band, and (b) in a p -type semiconductor, it is located between the acceptor energy level and the top of the valence band



semiconductor. This proximity also has the very important consequence that the concentration of holes in the valence band is much larger than for an intrinsic semiconductor, as a result of the shape of the Fermi-Dirac distribution shown in Fig. 5.12, at room temperature. And these holes can easily participate in electrical conduction phenomena.

For very high temperatures, all the donors or acceptors are ionized, and we have $N_D^+ = N_D$ or $N_A^+ = N_A$. Thus, the contribution from dopants to the charged carriers is limited, which is typically to a maximum of 10^{19} cm^{-3} . At the same time, the intrinsic contribution to the concentrations of electrons and holes, given by Eq. (7.32), is such that (take $T \rightarrow \infty$):

$$n = p = n_i = \sqrt{N_c N_v} \exp\left(-\frac{E_g}{2k_b T}\right) \approx \sqrt{N_c N_v} \tag{7.38}$$

Moreover, from Eqs. (7.19) and (7.28), we saw that the effective density of states N_c and N_v both increase as $T^{3/2}$. Therefore, the intrinsic contribution to n and p also increases as $T^{3/2}$, i.e., is not limited when the temperature increases, unlike the contribution from dopants. The charge neutrality relation in Eq. (7.33) then becomes:

$$n \approx p \tag{7.39}$$

This means that at very high temperatures, the charge carriers in an extrinsic semiconductor behave as in an intrinsic semiconductor. This also means that the Fermi energy tends to the expression given in Eq. (7.37). From these qualitative arguments, we can schematically illustrate the evolution of the Fermi energy as a function of temperature in Fig. 7.10 for an n -type and a p -type semiconductor.

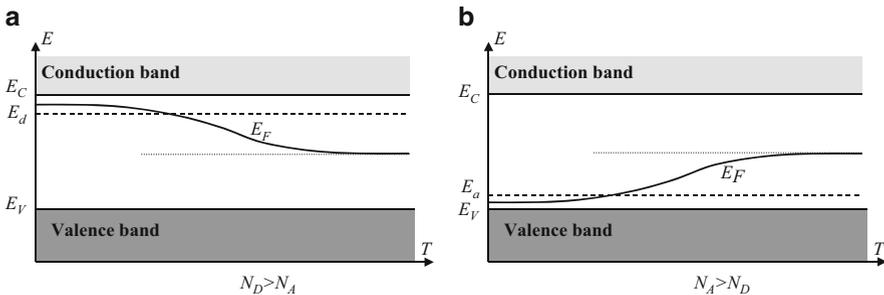


Fig. 7.10 Evolution of the Fermi energy as a function of temperature in a (a) n -type or (b) p -type semiconductor crystal. As the temperature is raised, the position of the Fermi energy shifts from its position in Fig. 7.9 to the position for an intrinsic semiconductor

7.9 Carrier Concentration in an n -Type Semiconductor

Before concluding this section on the electrical charge distribution at equilibrium, let us consider the example of a non-degenerate, n -type doped semiconductor. Here again, we will not go in a detailed numerical analysis but will provide the main qualitative results. The total dopant concentration will be denoted N_D . Assuming there is no acceptor ($N_A = 0$), the charge neutrality relation in Eq. (7.33) is now:

$$n = p + N_D^+ \quad (7.40)$$

Several levels of approximations, corresponding to several temperature regimes, can be considered to further simplify this expression. But before continuing the discussion, we should point out that holes in this semiconductor can only originate from the intrinsic contribution, not an extrinsic source such as a dopant (we chose $N_A = 0$).

The first regime corresponds to high temperatures. As discussed in the previous subsection, all the donors are ionized ($(N_D^+ = N_D)$). However, the concentrations of electrons n and holes p are much higher than the total concentration of donors ($n, p \gg N_D$), and they therefore obey the expressions derived for the intrinsic case, i.e.:

$$n \approx p \approx n_i = \sqrt{N_c N_v} \exp\left(-\frac{E_g}{2k_b T}\right) \quad (7.41)$$

As the temperature is lowered, while the donors remain ionized ($N_D^+ = N_D$), the intrinsic contribution to the concentrations of electrons and holes diminishes. Below a certain temperature, these contributions become negligible in comparison to N_D^+ or N_D . In this second temperature regime, p can be neglected ($p \ll N_D^+$) because the only contribution to p is the intrinsic contribution. Therefore, Eq. (7.40) becomes:

$$n \approx N_D \quad (7.42)$$

This is the most interesting characteristic of an extrinsic semiconductor. Indeed, if the concentration of donors can be intentionally controlled in the crystal during the synthesis, the concentration of electrons in the conduction band is precisely determined.

Specifically, the temperature at which the carrier concentration from thermal generation becomes equal to the background carrier concentration is called the intrinsic temperature T_i . Below T_i the carrier concentration is relatively temperature independent. Above T_i it increases exponentially with temperature.

As the temperature is further lowered, we reach a third regime where all the donors are not ionized anymore ($N_D^+ < N_D$). At the same time, we still have $p \ll N_D^+$. In this case, Eq. (7.40) becomes:

$$n \approx N_D^+ \quad (7.43)$$

At low temperatures, the Fermi energy E_F lies between the bottom of the conduction band E_C and the donor level E_d . Therefore, $E_F - E_d > 0$ and the expression for N_D^+ in Eq. (7.34) can be simplified to become:

$$\frac{N_D^+}{N_D} = \frac{1}{2\exp\left(\frac{E_F - E_d}{k_b T}\right) + 1} \approx \frac{1}{2\exp\left(\frac{E_F - E_d}{k_b T}\right)}$$

or:

$$N_D^+ \approx \frac{N_D}{2} \exp\left(-\frac{E_F - E_d}{k_b T}\right) \quad (7.44)$$

Let us now calculate the product nN_D^+ . On the one hand, it is equal to n^2 from Eq. (7.43). On the other hand, it is equal to:

$$N_c \exp\left(\frac{E_F - E_C}{k_b T}\right) \frac{N_D}{2} \exp\left(-\frac{E_F - E_d}{k_b T}\right) \quad (7.45)$$

after using Eqs. (7.21) and (7.44). We then obtain:

$$n^2 \approx \frac{N_c N_D}{2} \exp\left(-\frac{E_C - E_d}{k_b T}\right) \quad (7.46)$$

which yields:

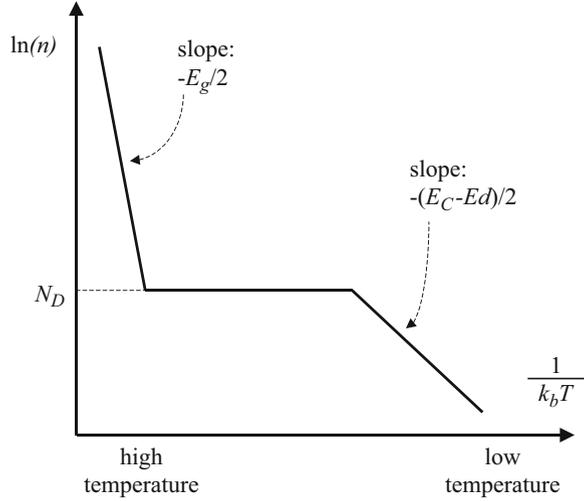
$$n \approx \sqrt{\frac{N_c N_D}{2}} \exp\left(-\frac{E_C - E_d}{2k_b T}\right) \quad (7.47)$$

The three expressions of n in Eqs. (7.41), (7.42), and (7.47) provide good approximations of the concentration of electrons in the conduction band as a function of temperature. It is customary to plot this concentration in a logarithmic scale for n and as a function of inverse temperature (i.e., $\frac{1}{k_b T}$), so that the slopes of the curve can be directly correlated to the bandgap energy E_g in Eq. (7.47) and the ionization energy $E_C - E_d$ in Eq. (7.47)). This is very simply shown in the schematic diagram in Fig. 7.11. Here, the temperature dependence of N_c ($T^{3/2}$) from Eq. (7.19)) has been neglected in comparison to the temperature dependence of the exponential terms.

In the case of a *p*-type semiconductor, with an acceptor concentration N_A , the following hole concentrations for the various regimes discussed previously can be determined.

In the first regime, at high temperatures, the concentrations of holes p and electrons n are much higher than the total concentration of acceptors ($n, p \gg N_A$) and thus follow their expressions for the intrinsic case, as in Eq. (7.41):

Fig. 7.11 Simple schematic diagram of the dependence of the electron concentration in the conduction band as a function of temperature in a typical n -type semiconductor crystal. At low temperatures, the carrier concentration follows a relation dependent on the donor energy inside the bandgap. At moderate temperatures, the electron concentration is nearly constant equal to the donor concentration. At high temperatures, the carrier concentration approaches that of an intrinsic semiconductor



$$n \approx p \approx n_i = \sqrt{N_c N_v} \exp\left(-\frac{E_g}{2k_b T}\right) \quad (7.48)$$

In the second regime, Eq. (7.42) can be transformed for a p -type semiconductor into:

$$p \approx N_A \quad (7.49)$$

In the third regime, as the temperature is further lowered, Eq. (7.43) can also be transformed for a p -type semiconductor into:

$$p \approx N_A^- \quad (7.50)$$

From Eq. (7.35), using the same derivation as between Eqs. (7.44) and (7.47), we get:

$$p \approx \sqrt{\frac{N_v N_D}{4}} \exp\left(-\frac{E_a - E_v}{2k_b T}\right) \quad (7.51)$$

7.10 Summary

In this chapter, we have first described the equilibrium properties of charge carriers in a semiconductor. We introduced the concepts of effective density of states, mass action law, and intrinsic and extrinsic semiconductor. The n -type and p -type doping of semiconductors has been discussed, taking into account the charge neutrality of the solid. We also discussed the importance of the Fermi energy.

Problems

1. Calculate the conduction band effective density of states for Si, Ge, and GaAs at 300 K. Plot it in logarithmic scale as a function of the logarithm of the temperature.
2. Calculate the valence band effective density of states for Si, Ge, and GaAs at 300 K. Plot it as a function of temperature, in logarithmic scale. We know that the valence band is degenerate at the center of the Brillouin zone as there is a heavy-hole band (with effective mass m_{hh}) and a light-hole band (with effective mass m_{lh}). The effective mass to be used in Eq. (7.28) is then:

$$m_h = \left(m_{hh}^{3/2} + m_{lh}^{3/2} \right)^{2/3}.$$

3. Find the energies at which the distribution of electrons in the conduction band and the distribution of holes in the valence band have maxima, if distributions are governed by Maxwell-Boltzmann statistics.
4. Estimate relative errors in the calculation of free carrier concentration when the Maxwell-Boltzmann statistics is applied for semiconductors with Fermi energy within the energy gap, if the Fermi level is $3k_bT$, $2k_bT$, and k_bT away from the bandgap edge or if it coincides with the edge. Use the given table of the exact value of the Fermi integral ($F_{1/2}$) for the comparison.
5. Calculate the intrinsic carrier concentrations for Si, Ge, GaAs, and GaN at 300 K, in the non-degenerate case. Plot their evolution as a function of temperature, in logarithmic scale.
6. From the periodic table, give examples of n -type and p -type dopants for Ge and GaAs. Is silicon an n -type or a p -type dopant in GaAs? Interpret.
7. As we know P is an n -type dopant for Si and Ge. Nitrogen is in the same column as P in the periodic table. Will N be a good dopant? Why?
8. Give an expression for the charge neutrality relation when double acceptors are present with a concentration N_{AA} . Double acceptors accept one or two electrons. Use the same notations as those in Sect. 7.3.
9. Plot the evolution of the Fermi energy as a function of temperature in intrinsic GaAs.
10. Consider a p -type doped GaAs semiconductor at 300 K with an experimentally measured hole concentration of $1.5 \times 10^{17} \text{ cm}^{-3}$. The p -type dopant has an energy level such that $\Delta E_a = E_a - E_V = 125 \text{ meV}$. Assuming there is no donor, determine the proportion of ionized acceptors. Determine the total concentration of acceptors.
11. Consider an n -type doped GaAs semiconductor at 300 K with an experimentally measured electron concentration of $3 \times 10^{17} \text{ cm}^{-3}$. The n -type dopant has an energy level such that $\Delta E_d = E_C - E_d = 25 \text{ meV}$. Assuming there is no acceptor, determine the proportion of ionized donors. Determine the total concentration of donors.

12. Derive expressions for concentrations of free carriers in a semiconductor doped with both, donor and acceptor impurities. Determine the conductivity type and calculate the concentrations of carriers in silicon at $T = 300$ K, if it is doped with:
- $N_A = 10^{16} \text{ cm}^{-3} \gg N_D$.
 - $N_D = 10^{16} \text{ cm}^{-3} \gg N_A$.
 - $N_D = N_A = 10^{16} \text{ cm}^{-3}$.
- Assume that all impurities are ionized and $n_i = 1.38 \times 10^{10} \text{ cm}^{-3}$ at 300 K.
13. Calculate the concentration of acceptor impurities in silicon, and determine the type of semiconductor, if at $T = 300$ K the concentration of electrons is $5 \times 10^{11} \text{ cm}^{-3}$ and the concentration of donor impurities is 10^{15} cm^{-3} . Assume $n_i = 1.38 \times 10^{10} \text{ cm}^{-3}$ at 300 K.
14. Calculate concentrations of carriers in silicon doped by acceptors $N_A = 10^{14} \text{ cm}^{-3}$ at:
- 27 °C
 - 175 °C

References

- Sze SM (1981) *Physics of Semiconductor Devices*. Wiley, New York
 Wolfe CM, Holonyak N Jr, Stillman GE (1989) *Physical properties of semiconductors*. Prentice-Hall, Englewood Cliffs

Further Reading

- Anselm A (1981) *Introduction to semiconductor theory*. Prentice-Hall, Englewood Cliffs
 Ashcroft NW, Mermin ND (1976) *Solid State Physics*. Holt, Rinehart and Winston, New York
 Cohen MM (1972) *Introduction to the quantum theory of semiconductors*. Gordon and Breach, New York
 Ferry DK (1991) *Semiconductors*. Macmillan, New York
 Hummel RE (1986) *Electronic properties of materials*. Springer, New York
 Pierret RF (1989) *Advanced semiconductor fundamentals*. Addison-Wesley, Reading
 Sapoval B, Hermann C (1995) *Physics of Semiconductors*. Springer, New York
 Streetman BG (1990) *Solid state electronic devices*. Prentice-Hall, Englewood Cliffs
 Wang S (1989) *Fundamentals of semiconductor theory and device physics*. Prentice-Hall, Englewood Cliffs