Chapter 18
Electrons and Holes in Semiconductors at Equilibrium

18.1 Introduction

Purpose of this chapter is providing the equilibrium expressions of the electron and hole concentrations in a semiconductor. For comparison, the cases of insulators and conductors are discussed qualitatively, and the concepts of conduction band, valence band, and generation of an electron-hole pair are introduced. The important issue of the temperature dependence of the concentrations is also discussed. Then, the general expressions of the concentrations in an intrinsic semiconductor are worked out, followed by an estimate of the Fermi level's position. Next, the equilibrium expressions are worked out again, this time in the case where substitutional impurities of the donor or acceptor type are present within the semiconductor. The mechanism by which donor-type dopants provide electrons to the conduction band, and acceptor-type dopants provide holes to the valence band, is explained. An important outcome of the analysis is that the introduction of suitable dopants makes the concentration of majority carriers practically independent of temperature, at least in a range of temperatures of practical interest for the functioning of integrated circuits. The simplifications due to the complete-ionization and non-degeneracy conditions are illustrated, along with the compensation effect. Finally, the theory is extended to the case of a non-uniform doping distribution, where the concentrations must be calculated self-consistently with the electric potential by solving the Poisson equation. The last section illustrates the band-gap narrowing phenomenon. In the complements, after a brief description of the relative importance of germanium, silicon, and gallium arsenide in the semiconductor industry, a qualitative analysis of the impurity levels is carried out by an extension of the Kronig–Penney model, and the calculation of the position of the impurity levels with respect to the band edges is carried out.

M. Rudan, *Physics of Semiconductor Devices,*
DOI 10.1007/978-1-4939-1151-6_18

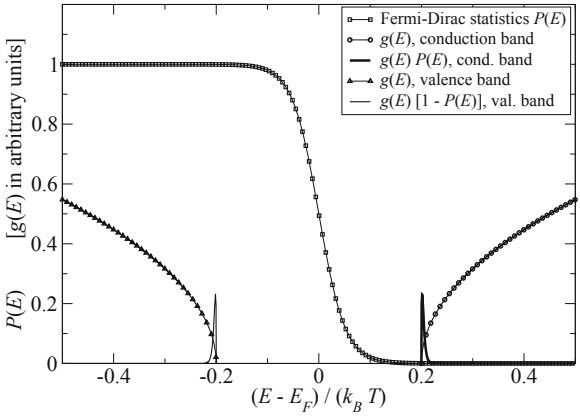

Fig. 18.1 Description of the particles' population in the conduction and valence bands of an insulator. To make them more visible, the products $g(E)\,P(E)$ and $g(E)\,[1-P(E)]$ have been amplified, with respect to $g(E)$ alone, by a factor 2×10^{31} (compare with Figs. 17.14 and 18.2). The gap's extension is arbitrary and does not refer to any specific material

18.2 Equilibrium Concentration of Electrons and Holes

The expressions worked out in this chapter are obtained by combining the information about the band structure of the crystal under investigation, given in Chap. 17, with that of the equilibrium distribution of fermions (Chap. 15). A qualitative description is given first, starting from the simplified case where the structures of the conduction and valence band are symmetric (Fig. 17.14).

Considering again a case where the Fermi level E_F coincides with the gap's midpoint and $M_V\,m_h^{3/2}=M_C\,m_e^{3/2}$, let the only difference with respect to the material of Fig. 17.14 be that the energy gap is larger (Fig. 18.1). Despite the fact that the situations illustrated in the two figures may appear similar to each other, it must be realized that a 2×10^{31} amplification factor, with respect to the scale of $g(E)$ alone, is necessary to make the products $g(E)\,P(E)$, $g(E)\,[1-P(E)]$ visible in Fig. 18.1, in contrast to the 10^3 factor used in Fig. 17.14. In practice, for the material of Fig. 18.1 the states in the conduction band are empty and those of the valence band are full.[1] Remembering that a band whose states are empty and, similarly, a band whose states are fully occupied, do not provide any conduction, it turns out that the electrical conductivity of a material like that of Fig. 18.1 is expected to vanish: the material is an *insulator*.

A different case is found when the Fermi level is inside a band, like in the crystal illustrated in Fig. 18.2. The name *conduction band* is given in this case to the band where the Fermi level belongs; the band beneath is called *valence band* also in this case. Due to the position of the Fermi level, the parabolic-band approximation is

[1] The curves of Figs. 17.14, 18.1 are drawn in arbitrary units. To better appreciate the difference in a practical situation, one may use the equilibrium concentration of electrons in silicon at $T=300$ K, which is about 10^{16} m^{-3}. Thus, if Fig. 17.14 is thought of as representing silicon, the ratio of the amplification factors used in Figs. 17.14 and 18.1 produces, in the latter, a concentration of about one electron in 1000 km^3.

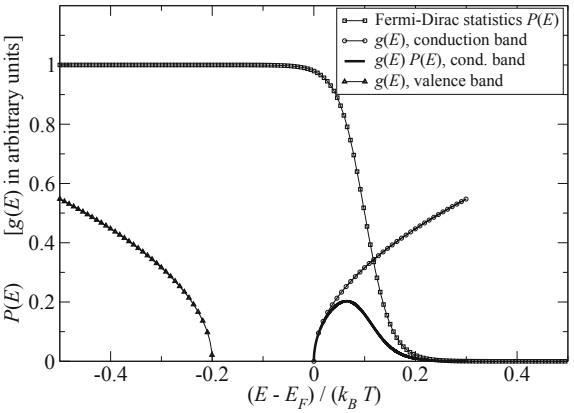

Fig. 18.2 Description of the electron population in the conduction band of a conductor. The product $g(E)\,P(E)$ is drawn in the same scale as $g(E)$ alone (compare with Figs. 17.14 and 18.1). The gap's extension is arbitrary and does not refer to any specific material

grossly mistaken, and is used for a qualitative discussion only. In the figure, the $g(E)\,P(E)$ product is drawn without using any amplification factor, this showing that the electron concentration in the conduction band is much larger than for a semiconductor; the latter band, in turn, is the only one that contributes to conduction, because the concentration of holes in the valence band is negligible. Due to the much larger concentration of charges one expects that the conductivity of the crystal under investigation be large; in fact, this is found to be the case, and the crystal is a *conductor*.

In summary, the combination of a few factors: position of the Fermi level with respect to the bands, gap's width, and structure of the density of states, dictates the presence of partially-filled bands and the concentration of charges in them. Other situations may occur besides those depicted in Figs. 17.14, 18.1, and 18.2, e.g., the Fermi level may be positioned within the gap, but closer to one of the band edges than to the other. They are illustrated in Sects. 18.4.1 and 18.4.2.

Another important issue is the dependence of the electron and hole populations on temperature. Considering the case of a semiconductor first, the discussion is carried out with reference to Figs. 17.14 and 15.3; the latter illustrates the temperature dependence of the Fermi-Dirac statistics, using the simplifying hypothesis that the position of the Fermi level does not change with temperature.[2] Following the reasoning carried out in Sect. 17.6.5, the temperature dependence of the effective masses is neglected as well. When temperature increases, the occupation probability of the conduction-band states increases, this making the concentration of electrons in the conduction band to increase; at the same time, the occupation probability of the valence-band states decreases, this making the concentration of holes in the valence band to increase as well. This outcome is easily understood if one thinks that the increase in temperature is achieved by transferring energy from an external reservoir

[2] The temperature dependence of the Fermi level is influenced by the shape of the density of states [54]. Such a dependence can be neglected in the qualitative discussion carried out here.

to the semiconductor; part of this energy is absorbed by the nuclei and produces a change in the equilibrium distribution of phonons (Sect. 16.6), while the other part is absorbed by the electrons and produces a redistribution of the latter within the available energy states. The absorption of energy by electrons that, prior to the temperature increase, belonged to valence-band states, makes some of them to transit to the conduction band. As each electron that makes a transition leaves a hole in the valence band, this is an example of *generation of an electron–hole pair* (compare with Sect. 17.6.6).

As shown in Sect. 19.5.5, the conductivity of a semiconductor is proportional to $\mu_n\, n + \mu_p\, p$, where n is the concentration of conduction-band electrons, p the concentration of valence-band holes, and μ_n, μ_p the electron and hole mobilities, respectively. Mobilities account for the scattering events undergone by electrons and holes during their motion within the material, and are found to decrease when temperature increases. It follows that the decrease in mobility competes with the increase of the concentrations in determining the temperature dependence of conductivity in a semiconductor. In practice, the increase in the concentrations is much stronger due to the exponential form of the Fermi-Dirac statistics, so that the conductivity of a semiconductor strongly increases with temperature.[3]

The qualitative analysis of conductivity is the same for a conductor where, as holes are absent, the conductivity is proportional to $\mu_n\, n$. However, the outcome is different; in fact, while μ_n, like that of a semiconductor, decreases when temperature increases, the electron concentration n is unaffected by temperature. In fact, from Fig. 18.2 one finds that the deformation of the Fermi-Dirac statistics due to a temperature variation produces a rearrangement of the electron distribution within the conduction band itself, and no electron-hole generation; as a consequence, the energies of some electrons of the conduction band change, while the electron number (hence the concentration n) does not. As mobility depends weakly on temperature, the conductivity of a conductor turns out to slightly decrease as temperature increases.

The calculation of the equilibrium electron concentration in the conduction band of a semiconductor is based on (17.76), where the density of states in energy g is replaced with the combined density of states in energy and volume γ given by (17.73). The concentration of electrons and the Fermi level are indicated here with n_i, E_{Fi} instead of n, E_F to remark the fact that the semiconductor is free from impurities (for this reason, the semiconductor is called *intrinsic*). From the discussion of Sect. 17.6.5, the parabolic-band approximation is acceptable, so that the combined density of states is given by (17.73).[4] Remembering that the lower edge of the conduction band is indicated with E_C, and letting E_{CU} be the upper edge of the same band, one finds

$$n_i = \int_{E_C}^{E_{CU}} \gamma(E)\, P(E)\, \mathrm{d}E \simeq \int_{E_C}^{+\infty} \frac{\sqrt{2}\, M_C\, m_e^{3/2}\, \sqrt{E - E_C}/(\pi^2\, \hbar^3)}{\exp\left[(E - E_{Fi})/(k_B\, T)\right] + 1}\, \mathrm{d}E, \quad (18.1)$$

[3] This is a negative aspect because the electrical properties of the material are strongly influenced by the ambient temperature. The drawback is absent in doped semiconductors, at least in the range of temperatures that are typical of the operating conditions of semiconductor devices (Sect. 18.4.1).

[4] As before, the band index is dropped from (17.73).

where the upper integration limit has been replaced with $+\infty$ on account of the fact that the integrand vanishes exponentially at high energies. Using the auxiliary variables

$$\zeta_e = E_C - E_{Fi}, \qquad x = \frac{E - E_C}{k_B\, T} \geq 0, \qquad \xi_e = -\frac{\zeta_e}{k_B\, T}, \tag{18.2}$$

with x, ξ_e dimensionless, transforms (18.1) into

$$n_i = \frac{\sqrt{2}}{\pi^2 \hbar^3}\, M_C m_e^{3/2}\, (k_B\, T)^{3/2} \int_0^{+\infty} \frac{\sqrt{x}}{\exp\,(x - \xi_e) + 1}\, \mathrm{d}x. \tag{18.3}$$

From the definition (C.109) of the Fermi integral of order 1/2 it then follows

$$n_i = N_C\, \Phi_{1/2}(\xi_e), \qquad N_C = 2\, M_C \left(\frac{m_e}{2\,\pi\,\hbar^2}\, k_B\, T\right)^{3/2}, \tag{18.4}$$

with N_C the *effective density of states* of the conduction band. Observing that $\Phi_{1/2}$ is dimensionless, the units of N_C are m^{-3}.

The concentration of holes in the valence band is determined in a similar manner, namely, starting from an integral of the form (17.76), where $P(E)$ is replaced with $1 - P(E)$ and the density of states in energy g is replaced with the combined density of states in energy and volume γ. The concentration of holes is indicated here with p_i and, as for the electrons, the parabolic-band approximation is adopted, so that the combined density of states is obtained from (17.74). Finally, $1 - P(E)$ is expressed through (17.77). Remembering that the upper edge of the valence band is indicated with E_V, and letting E_{VL} be the lower edge of the same band, one finds

$$p_i = \int_{E_{VL}}^{E_v} \gamma(E)\,[1 - P(E)]\,\mathrm{d}E \simeq \int_{-\infty}^{E_V} \frac{\sqrt{2}\, M_V\, m_h^{3/2}\, \sqrt{E_V - E}/(\pi^2\,\hbar^3)}{\exp\,[(E_{Fi} - E)/(k_B\, T)] + 1}\, \mathrm{d}E, \tag{18.5}$$

where the lower integration limit has been replaced with $-\infty$ on account of the fact that the integrand vanishes exponentially at low energies. Using the auxiliary variables

$$\zeta_h = E_{Fi} - E_V, \qquad x = \frac{E_V - E}{k_B\, T} \geq 0, \qquad \xi_h = -\frac{\zeta_h}{k_B\, T}, \tag{18.6}$$

transforms (18.5) into

$$p_i = \frac{\sqrt{2}}{\pi^2 \hbar^3}\, M_V m_h^{3/2}\, (k_B\, T)^{3/2} \int_0^{+\infty} \frac{\sqrt{x}}{\exp\,(x - \xi_h) + 1}\, \mathrm{d}x, \tag{18.7}$$

whence, introducing the *effective density of states* of the valence band, N_V,

$$p_i = N_V\, \Phi_{1/2}(\xi_h), \qquad N_V = 2\, M_V \left(\frac{m_h}{2\,\pi\,\hbar^2}\, k_B\, T\right)^{3/2}. \tag{18.8}$$

Table 18.1 Gap and average effective masses of silicon, germanium, and gallium arsenide[a]

Material	E_{G0} (eV)	α (eV/K)	β (K)	$E_G(T_a)$	m_e/m_0	m_h/m_0
Si	1.160	7.02×10^{-4}	1108	1.12	0.33	0.56
Ge	0.741	4.56×10^{-4}	210	0.66	0.22	0.31
GaAs	1.522	5.80×10^{-4}	300	1.43	0.68	0.50

[a]Symbol T_a indicates the room temperature

18.3 Intrinsic Concentration

Equations (18.2, 18.4) and (18.6, 18.8) express the equilibrium concentrations of electron and holes in a semiconductor in terms of temperature and of the distance of the Fermi level from the edge E_C of the conduction band or, respectively, from the edge E_V of the valence band. Obiouvsly the two distances are not independent from each other; from (18.2, 18.6) one finds in fact

$$-(\xi_e + \xi_h) = \frac{\zeta_e + \zeta_h}{k_B\, T} = \frac{E_C - E_{Fi} + E_{Fi} - E_V}{k_B\, T} = \frac{E_G}{k_B\, T}, \tag{18.9}$$

where $E_G = E_C - E_V$ is the extension of the semiconductor's gap, known from the calculation of the band structure, or also from electrical or optical measurements. As the band structure is influenced by temperature, the gap depends on temperature as well (Sect. 17.6.5); such a dependence is important because it strongly influences the concentration of electrons and holes. The results of gap's calculations or measurements, that show that E_G decreases when temperature increases, are usually rendered in compact form by means of interpolating expressions, an example of which is

$$E_G(T) \approx E_{G0} - \alpha \frac{T^2}{T + \beta}, \tag{18.10}$$

where E_{G0} is the gap's width extrapolated at $T = 0$ and $\alpha > 0$, $\beta > 0$ are material's parameters. Table 18.1 reports the parameters related to the gap's width for Si, Ge, and GaAs, along with the values of the average effective masses normalized to the rest mass of the free electron [103, Chap. 2–3]. The plot of $E_G(T)$ is shown in Fig. 18.3 for the three semiconductors of Table 18.1.

Note that expressions (18.4, 18.8) can be used only if the position of the Fermi level is known; in fact, the latter (which is unknown as yet) enters the definitions (18.2, 18.6) of parameters ξ_e, ξ_h. To proceed one remembers that in the $T \to 0$ limit the Fermi-Dirac statistics becomes discontinuous, specifically it is $P = 1$ for $E < E_{Fi}$ and $P = 0$ for $E > E_{Fi}$ (Sect. 15.8.1); on the other hand, the experimental evidence shows that in the $T \to 0$ limit the conductivity of a semiconductor vanishes: this corresponds to a situation where all states of the conduction band are empty while those of the valence bands are filled. In conclusion, when $T \to 0$ the Fermi level is still positioned in the gap and it is $n_i = p_i = 0$. If, starting from this situation, the temperature is brought again to some finite value $T > 0$, such that some of the

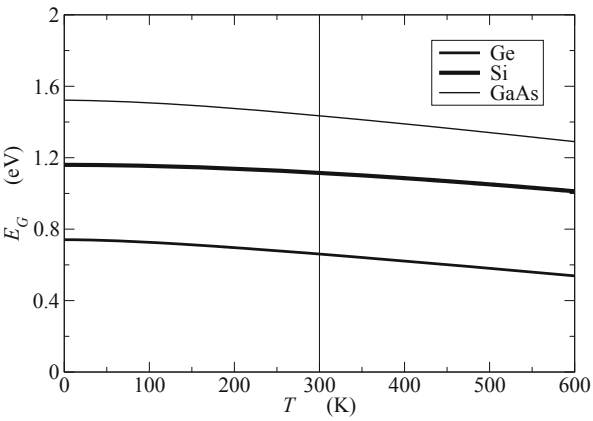

Fig. 18.3 Plot of the gap as a function of temperature for Ge, Si, and GaAs. The *vertical line* marks $T = 300$ K

valence-band electrons transit to the conduction band, the total number of holes thus formed equals that of the transited electrons. Due to the spatial uniformity of the material, the concentrations are equal to each other as well; in conclusion it is[5]

$$n_i = p_i, \qquad M_C\, m_e^{3/2}\, \Phi_{1/2}(\xi_e) = M_V\, m_h^{3/2}\, \Phi_{1/2}(\xi_h), \tag{18.11}$$

the second of which has been obtained by deleting the common factors from (18.4) and (18.8). Relations (18.9) and (18.11) are a set of two equations in the unknowns ξ_e, ξ_h, whose solution allows one to determine the position of the Fermi level through (18.2) or (18.6).

The second relation in (18.11) can also be exploited for carrying out an estimate of ξ_e, ξ_h, basing on the values of the masses given in Table 18.1. To this purpose one observes[6] that the ratio $M_C\, m_e^{3/2}/(M_V\, m_h^{3/2})$ is about 1.4, 1.2, 0.8 for Si, Ge, and GaAs, respectively, so that a crude estimate is obtained by letting $\Phi_{1/2}(\xi_e) \simeq \Phi_{1/2}(\xi_h)$. As the Fermi integral is a monotonic function of the argument (Sect. C.13), it follows $\xi_e \simeq \xi_h$. Replacing from (18.2, 18.6) yields $E_{Fi} \simeq (E_C + E_V)/2$, and $\zeta_e \simeq \zeta_h \simeq E_G/2 > 0$.

The usefulness of this estimate actually lies in that it simplifies the Fermi integrals. Taking for instance the case of room temperature, it is $k_B\, T_a \simeq 26$ meV; using the values of $E_G(T_a)$ from Table 18.1 shows that $E_G \gg k_B T_a$, whence $-\xi_e \gg 1$ and $-\xi_h \gg 1$, so that the approximate expression (C.105) applies: $\Phi_{1/2}(\xi) \simeq \exp(\xi)$. In conclusion, (18.4, 18.8) simplify to $n_i \simeq N_C\, \exp(\xi_e)$, $p_i \simeq N_V\, \exp(\xi_h)$, namely,

$$n_i \simeq N_C\, \exp\left(-\frac{E_C - E_{Fi}}{k_B\, T}\right), \qquad p_i \simeq N_V\, \exp\left(-\frac{E_{Fi} - E_V}{k_B\, T}\right). \tag{18.12}$$

[5] Note that the reasoning leading to the first relation in (18.11) is not limited to the case of the parabolic-band approximation, but holds for a general form of the densities of states.

[6] Remembering the discussion carried out in Sect. 17.6.5 it is $M_C(\text{Si}) = 6$, $M_C(\text{Ge}) = 4$, $M_C(\text{GaAs}) = 1$, $M_V = 2$.

As the two concentrations are equal to each other it is customary to use the same symbol n_i for both; the product of the two expressions (18.12) combined with (18.9) yields

$$n_i\, p_i = n_i^2 \simeq N_C\, N_V\, \exp\left(\xi_e + \xi_h\right) = N_C\, N_V\, \exp\left[-E_G/(k_B\, T)\right]. \tag{18.13}$$

The expression of the *intrinsic concentration* thus reads

$$n_i \simeq \sqrt{N_C\, N_V}\, \exp\left(-\frac{E_G}{2\, k_B\, T}\right), \qquad N_C\, N_V = \frac{M_C\, M_V}{2\, \pi^3\, \hbar^6}\, (m_e\, m_h)^{3/2}\, (k_B\, T)^3\,. \tag{18.14}$$

Its values at room temperature are listed in Table 18.2 along with those of the effective densities of states, for Si, Ge, and GaAs. As for the temperature dependence of n_i one notes that, besides appearing in the exponent's denominator in the first relation of (18.14), the lattice temperature also influences the numerator E_G and the $N_C\, N_V$ factor. Among these dependencies, that of the exponent's denominator is by far the strongest; it follows that a first-hand description of $n_i(T)$ can be given by considering only the latter. This yields the Arrhenius plots shown in Fig. 18.4, where the relations $n_i(T)$ reduce to straight lines. It is important to note that, despite the similarity of the effective densities of states for the three semiconductors considered here (Table 18.2), the intrinsic concentrations differ by orders of magnitude. This is due to the exponential dependence of n_i on E_G, which amplifies the differences in E_G (visible in Table 18.1) of the three semiconductors.

The estimate of the position of the Fermi level in an intrinsic semiconductor carried out above has led to the conclusion that the Fermi integrals used for calculating the intrinsic concentrations can be replaced with exponentials.[7] Such a conclusion can now be exploited for a more precise calculation of the Fermi level's position, where the coefficients of (18.11) are kept. Using the exponentials in (18.11) one finds

$$M_C\, m_e^{3/2}\, \exp\left(\xi_e\right) = M_V\, m_h^{3/2}\, \exp\left(\xi_h\right). \tag{18.15}$$

Taking the logarithm of both sides and using (18.9) yields $\xi_h - \xi_e = \xi_h + \xi_e - 2\,\xi_e = \log\left[M_C\, m_e^{3/2}/(M_V\, m_h^{3/2})\right]$, whence

$$E_C - E_{Fi} = \frac{E_G}{2} + \frac{k_B\, T}{2}\, \log\left(\frac{M_C\, m_e^{3/2}}{M_V\, m_h^{3/2}}\right). \tag{18.16}$$

The second term at the right hand side of (18.16) is the correction with respect to the estimate carried out earlier. In the $T \to 0$ limit, the Fermi level in the intrinsic semiconductors under consideration coincides with the gap's midpoint. When the temperature increases, if $M_C\, m_e^{3/2} > M_V\, m_h^{3/2}$ the distance $E_C - E_{Fi}$ becomes

[7] This is not necessarily true for an *extrinsic* semiconductor, where suitable impurity atoms are introduced into the semiconductor lattice (Sects. 18.4.1, 18.4.2).

Table 18.2 Intrinsic concentrations of silicon, germanium, and gallium arsenide[a]

Material	$N_C(T_a)$ (cm^{-3})	$N_V(T_a)$ (cm^{-3})	$n_i(T_a)$ (cm^{-3})
Si	2.82×10^{19}	1.05×10^{19}	7.61×10^{9}
Ge	1.04×10^{19}	0.43×10^{19}	1.39×10^{12}
GaAs	4.45×10^{19}	0.99×10^{19}	2.40×10^{6}

[a]In this field it is customary to express the concentrations in cm^{-3} instead of m^{-3}

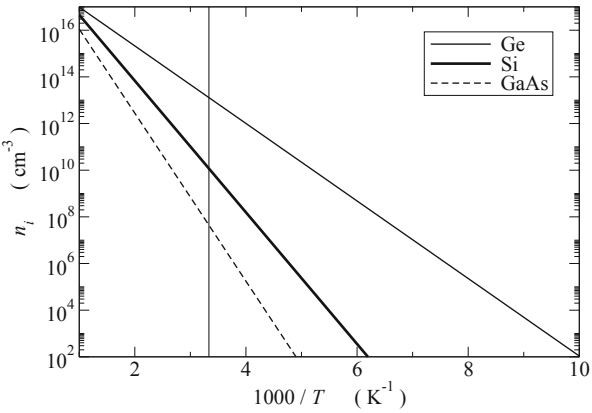

Fig. 18.4 Arrhenius plot of the intrinsic concentration in Ge, Si, and GaAs. The *vertical line* marks $T = 300$ K

larger, that is, the Fermi level moves towards the valence band; the opposite happens if $M_C\, m_e^{3/2} < M_V\, m_h^{3/2}$. For all practical purposes, considering that the argument of the logarithm in (18.16) is close to unity and the coefficient $k_B\, T$ is always small with respect to E_G, the position of the Fermi level can be thought of as coinciding with the gap's midpoint.

18.4 Uniform Distribution of Impurities

As described in Chap. 23, the fabrication of integrated circuits (IC) requires the introduction into the semiconductor material of atoms (called *impurities* or *dopants*) belonging to specifically-selected chemical species. Dopants are divided into two classes, termed *n-type* and *p-type*. With reference to silicon (Si), the typical *n*-type dopants are phosphorus (P), arsenic (As), and antimony (Sb), while the typical *p*-type dopants are boron (B), aluminum (Al), gallium (Ga), and Indium (In). For a qualitative introduction to the effects of dopants it is instructive to start with the case of an intrinsic semiconductor; the analysis is based on the simplified picture in which the original arrangement of the atoms in space, like that shown in Fig. 17.11, is deformed in such a way as to become two-dimensional. This representation, shown in Fig. 18.5 with reference to silicon, is convenient for the description carried out below. Each silicon atom has four electrons in the external shell, so that it can form

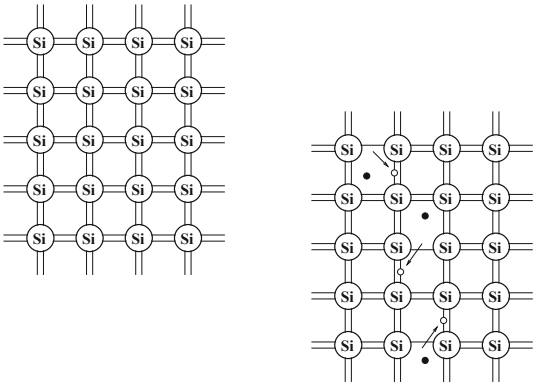

Fig. 18.5 Two-dimensional representation of the intrinsic silicon lattice. The *upper-left* part of the figure shows the $T \to 0$ limit

four covalent bonds[8] with other identical atoms; each pair of lines connecting two atoms in Fig. 18.5 stands for a pair of shared electrons. In the $T \to 0$ limit, the electrons are permanently bound to the atoms because the energy necessary to ionize is not available; this situation is drawn in the upper-left part of the figure. If, instead, the temperature is brought to some finite value $T > 0$ by transferring energy from an external reservoir to the semiconductor, part of this energy is absorbed by some of the electrons; the latter break the bond and become free to move within the material.[9] This situation is depicted in the lower-right part of Fig. 18.5, where the free electrons are represented with black dots.

When an electron becomes free and departs from an atom, the unbalanced positive charge left behind in the nucleus deforms the shape of the potential energy in the vicinity of the nucleus itself. The deformation is such that the potential-energy barrier that separates the nucleus from the neighboring ones becomes lower and thinner; this, in turn, enhances the probability of tunneling across the barrier by an electron belonging to a shared pair. Such a tunneling event restores the original pair, but leaves an unbalanced positive charge behind; instead of considering it as the motion of an electron from a complete to an incomplete pair, the tunneling event is more conveniently described as the opposite motion of an empty state, that is, a hole. This is indicated in Fig. 18.5 by the combinations of arrows and white dots.

The above description, based on a spatial picture of the material, is able to provide a qualitative explanation of the existence of electrons and holes in a semiconductor,[10] and completes the description given in Sect. 18.3, that focuses on an energy picture. It also constitutes the basis for analyzing the case of a doped semiconductor, as shown in the following sections.

[8] In a *covalent* bond atoms share their outermost electrons.

[9] Remembering that an equilibrium situation is considered here, the contributions to the electric current of these electrons cancel each other.

[10] A similar description could as well apply to a conductor. However, in such as case the barrier deformation is small and tunneling does not occur.

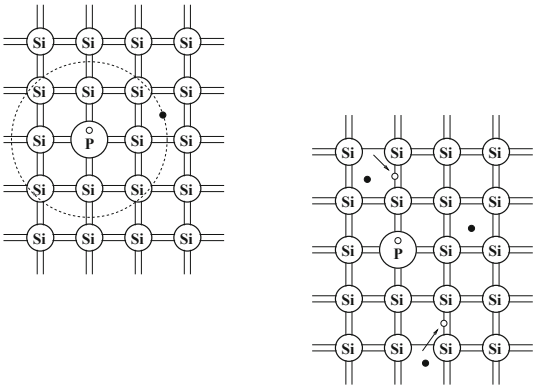

Fig. 18.6 Two-dimensional representation of the n-doped silicon lattice. The *upper-left* part of the figure shows the $T \to 0$ limit

18.4.1 Donor-Type Impurities

As shown in Chap. 23, when a dopant atom is introduced into the semiconductor lattice, in order to properly act as a dopant it must replace an atom of the semiconductor, namely, it must occupy a lattice position (*substitutional impurity*). The concentration of the dopant atoms that are introduced into a semiconductor is smaller by orders of magnitude than the concentration of the semiconductor atoms themselves. As a consequence, the average distance between dopant atoms within the lattice is much larger than that between the semiconductor atoms. Due to this, when the band structure is calculated, the modification in the potential energy introduced by the dopant atoms can be considered as a perturbation with respect to the periodic potential energy of the lattice; the resulting band structure is therefore the superposition of that of the intrinsic semiconductor and of a set of additional states, whose characteristics will be described later. For the moment being, the spatial description is considered, still using the two-dimensional picture, where it is assumed that phosphorus is used as dopant material (Fig. 18.6).

As the dopant concentration is small with respect to the semiconductor's, each phosphorus atom is surrounded by silicon atoms. Phosphorus has five electrons in the external shell: thus, it forms four covalent bonds with silicon, while the remaining electron does not participate in any bond. In the $T \to 0$ limit, the electrons are permanently bound to the atoms as in the intrinsic case; this situation is drawn in the upper-left part of the figure, where the electron that does not form any bond is represented, in a particle-like picture, as orbiting around the phosphorus atom. The orbit's radius is relatively large because the binding force is weak.[11] The white dot inside the phosphorus atom indicates the positive nuclear charge that balances the orbiting electron. If, instead, the temperature is brought to some finite value $T > 0$ by transferring energy from an external reservoir to the semiconductor, a fraction of

[11] In a more precise, quantum-mechanical description the electron is described as a stationary wave function extending over several lattice cells.

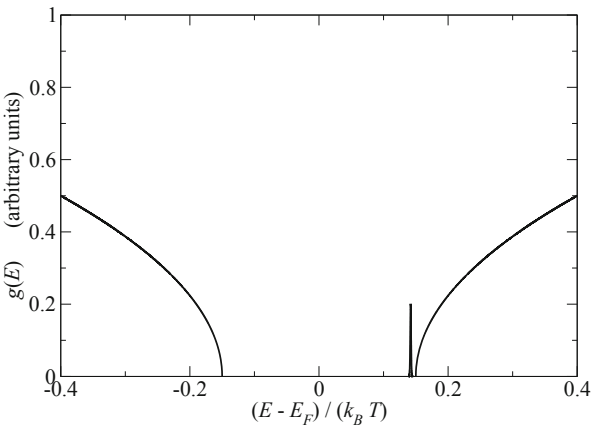

Fig. 18.7 Density of states in an *n*-doped semiconductor. The gap's extension is arbitrary and does not refer to any specific material

the orbiting electrons break the bond and become free to move within the material. At the same time, electrons belonging to shared pairs may also break their bonds, like in intrinsic silicon. This situation is depicted in the lower-right part of Fig. 18.6, where the free electrons are represented with black dots.

The deformation in the shape of the potential energy in the vicinity of a phosphorus nucleus is different from that near a silicon nucleus. In the latter case, the phenomenon is the same as in the intrinsic material: a series of tunneling events takes place, leading to the motion of holes from one site to a neighboring one. This is still represented by the combinations of arrows and white dots in Fig. 18.6. In the case of phosphorus, instead, the barrier deformation is small and tunneling does not occur: thus, the phosphorus atoms provide free electrons to the lattice, but no holes.

It is intuitive that the insertion of a prescribed amount of *n*-type dopants into the semiconductor lattice provides a method for controlling the number of free electrons and, through them, the material's conductivity. Moreover, if the ionization of the dopant atoms is not influenced by temperature, at least in a range of temperatures of practical interest, and the number of electrons made available by the dopant is dominant with respect to that provided by the intrinsic semiconductor, the drawback mentioned in Sect. 18.2 is eliminated: conductivity becomes temperature independent and its value is controlled by the fabrication process. This analysis is better specified below, starting from the consideration of the density of states in an *n*-type semiconductor, shown in Fig. 18.7.

18.4.1.1 Concentration of Ionized Impurities (Donor Type)

Atoms like that of phosphorus contribute electrons to the lattice; for this reason they are also called *donors*. The concentration of donors is indicated with N_D and, in this section, is assumed to be uniform in space. When donor atoms are present, the equilibrium concentrations of electrons and holes are different from those of the

intrinsic case; they are indicated with n and p, respectively, and are termed *extrinsic concentrations*. Their derivation is identical to that of the intrisic case and yields

$$n = N_C\, \Phi_{1/2}(\xi_e), \qquad p = N_V\, \Phi_{1/2}(\xi_h), \tag{18.17}$$

with N_C, N_V given by (18.4, 18.8), respectively, and

$$\xi_e = -\frac{\zeta_e}{k_B\, T} = -\frac{E_C - E_F}{k_B\, T}, \qquad \xi_h = -\frac{\zeta_h}{k_B\, T} = -\frac{E_F - E_V}{k_B\, T}. \tag{18.18}$$

The above expression of ξ_e, ξ_h are similar to (18.2, 18.6), the only difference being that the intrinsic Fermi level E_{Fi} is replaced here with the *extrinsic Fermi level* E_F.

In addition to the electrons of the conduction band and holes of the valence band, a third population of particles must be accounted for, namely, that associated to the dopants. Let $N_D^+ \leq N_D$ be the spatially-constant concentration of the ionized donors.[12] The difference $n_D = N_D - N_D^+$ is the concentration of the donor atoms that have not released the orbiting electron to the lattice. Equivalently, n_D may be thought of as the concentration of orbiting electrons that have not been released; such a concentration is given by

$$n_D = \int_{\Delta E_D} \gamma_D(E)\, P_D(E)\, \mathrm{d}E, \tag{18.19}$$

where $\gamma_D(E)$ is the combined density of states produced by the donor atoms, ΔE_D the energy range where $\gamma_D(E) \neq 0$, and $P_D(E)$ the occupation probability of such states. The form of γ_D depends on the concentration N_D; for low or moderately-high values of the donor concentration, γ_D has the form shown in Fig. 18.7. Such a density of states can be approximated by a Dirac delta centered at an energy E_D, called *donor level*, positioned within the gap at a small distance from the edge E_C of the conduction band:

$$\gamma_D(E) \simeq N_D\, \delta(E - E_D). \tag{18.20}$$

The coefficient in (18.20) is such that (18.19) yields $n_D = N_D$ when $P_D = 1$. The distance of the donor level from the minimum of the conduction band is calculated in Sect. 18.7.3; for the typical n-type dopants used with silicon it is $E_C - E_D \simeq 0.033\,\mathrm{eV}$. Another important feature of the donor levels (still in the case of low or moderately-high values of the impurity concentration) is that they are localized in space, as schematically illustrated in Fig. 18.8. For this reason, the Fermi-Dirac statistics describing the equilibrium distribution of the electrons within the donor states is slightly different from that used for the band electrons, and reads

[12] A phosphorous atom that has not released the orbiting electron is electrically neutral; when the orbiting electron breaks the bond and becomes free to move within the lattice, the atom becomes positively ionized. The symbol N_D^+ reminds one of that. A second ionization does not occur because the energy necessary for it is too high.

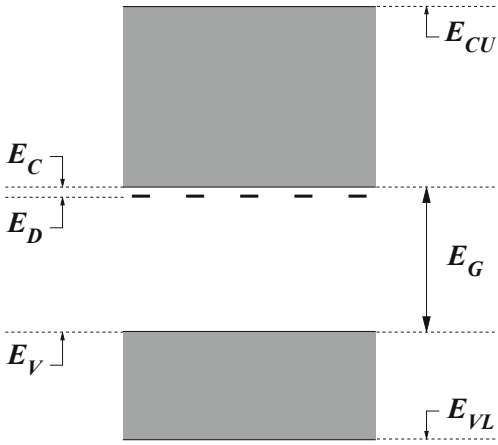

Fig. 18.8 Schematic representation of the donor states

$$P_D(E) = \frac{1}{(1/d_D)\,\exp\left[(E - E_F)/(k_B\,T)\right] + 1}, \tag{18.21}$$

with d_D the *donors' degeneracy coefficient* [23, Sect. 3.4]. In silicon, germanium, and gallium arsenide it is $d_D = 2$ [103, Sect. 2.4]. Combining (18.19) with (18.20) and (18.21) yields $n_D = N_D\,P_D(E_D)$, namely,

$$N_D^+ = N_D\,[1 - P_D(E_D)] = \frac{N_D}{d_D\,\exp\left[(E_F - E_D)/(k_B\,T)\right] + 1}. \tag{18.22}$$

Like in the intrinsic case discussed in Sect. 18.3, expressions (18.17, 18.22) can be used only if the position of the Fermi level is known. To proceed one considers again the $T \to 0$ limit, where the Fermi-Dirac statistics becomes discontinuous and the experimental evidence shows that the conductivity of a doped semiconductor vanishes: this corresponds to a situation where all states of the conduction band are empty, those of the valence bands are filled, and the donor states are filled as well (in other terms, no dopant atoms are ionized). In conclusion, when $T \to 0$ the Fermi level of an n-doped semiconductor is positioned between E_D and E_C, so that $n = p = N_D^+ = 0$. If, starting from this situation, the temperature is brought again to some finite value $T > 0$, such that some of the valence-band electrons and some of the electrons belonging to the dopant atoms transit to the conduction band, the total number of holes and ionized donors thus formed is equal to that of the transited electrons. Due to the spatial uniformity of the material, the same relation holds among the concentrations, so that

$$n = p + N_D^+, \tag{18.23}$$

whose limit for $N_D \to 0$ yields (18.11) as should be. Now, inserting (18.17, 18.22) into (18.23) after letting $\xi_D = (E_D - E_C)/(k_B\,T) < 0$, provides

$$N_C\,\Phi_{1/2}(\xi_e) = N_V\,\Phi_{1/2}(\xi_h) + \frac{N_D}{d_D\,\exp\left(\xi_e - \xi_D\right) + 1}. \tag{18.24}$$

The latter, along with the relation $\xi_e + \xi_h = -E_G/(k_B\,T)$, form a system in the two unknowns ξ_e, ξ_h, whose solution determines the position of E_F with respect to the band edges E_C and E_V at a given temperature $T > 0$ and donor concentration N_D. It is easily found that, for a fixed temperature, the argument $\xi_e = (E_F - E_C)/(k_B\,T)$ increases when N_D increases. In fact, using the short-hand notation $f(\xi_e) = d_D\,\exp(\xi_e - \xi_D) + 1 > 1$ one finds from (18.24)

$$\frac{\mathrm{d}N_D}{\mathrm{d}\xi_e} = N_D\,\frac{f-1}{f} + f\left[N_C\,\frac{\mathrm{d}\Phi_{1/2}(\xi_e)}{\mathrm{d}\xi_e} - N_V\,\frac{\mathrm{d}\Phi_{1/2}(\xi_h)}{\mathrm{d}\xi_h}\,\frac{\mathrm{d}\xi_h}{\mathrm{d}\xi_e}\right], \tag{18.25}$$

where the right hand side is positive because $\mathrm{d}\xi_h/\mathrm{d}\xi_e = -1$ and $\Phi_{1/2}$ is a monotonically-increasing function of the argument (Sect. C.13). In the intrinsic case $N_D = 0$ it is $E_F = E_{Fi}$ and $\xi_e < 0$; as N_D increases, the Fermi level moves towards E_C, this making ξ_e less and less negative. At extremely high concentrations of the donor atoms, the Fermi level may reach E_C and even enter the conduction band, this making ξ_e positive. As n, p, and N_D^+ are non negative, it is intuitive that an increasing concentration of donor atoms makes n larger than p; for this reason, in an n-doped semiconductor electrons are called *majority carriers* while holes are called *minority carriers*.[13]

The behavior of the Fermi level with respect to variations in temperature, at a fixed dopant concentration, can not be discussed as easily as that with respect to the dopant variations, because (18.24) is not expressible in the form $T = T(E_F)$ or its inverse. The analytical approach is made easier when some approximations are introduced into (18.24), as shown below. From a qualitative standpoint, one may observe that at very high temperatures the concentration of electron–hole pairs generated by the semiconductor prevails over the concentration of electrons provided by the dopant atoms, so that the position of the Fermi level must be close to the intrinsic one; in contrast, when $T \to 0$ the Fermi level is positioned between E_D and E_C as remarked above. In conclusion, for an n-doped semiconductor one expects that $\mathrm{d}E_F/\mathrm{d}T < 0$.

In view of further elaborations it is convenient to associate to the Fermi level an electric potential φ_F, called *Fermi potential*, defined by

$$q\,\varphi_F = E_{Fi} - E_F, \tag{18.26}$$

with $q > 0$ the elementary charge. In an n-type semiconductor it is $E_F > E_{Fi}$; thus, the Fermi potential is negative.

18.4.1.2 Complete Ionization and Non-Degenerate Condition (Donor Type)

When the Fermi level's position and temperature are such that

$$\xi_e - \xi_D < -1, \qquad E_F < E_D - k_B\,T, \tag{18.27}$$

[13] The electrons of the conduction band and the holes of the valence band are collectively indicated as *carriers*.

the exponential term at the right hand side of (18.22) is negligible with respect to unity, so that $N_D^+ \simeq N_D$. This condition is indicated with *complete ionization*. Note that, since $\xi_D < 0$, the complete-ionization condition in an n-doped semiconductor also implies $\xi_e < -1$; as a consequence, the approximation $\Phi_{1/2}(\xi_e) \simeq \exp(\xi_e)$ for the Fermi integral holds (Sect. C.13). Moreover, it is $\xi_h < \xi_e < -1$ because the Fermi level belongs to the upper half of the gap; thus, the same approximation applies to $\Phi_{1/2}(\xi_h)$ as well. When both equilibrium concentrations n and p are expressible through the approximation $\Phi_{1/2}(\xi) \simeq \exp(\xi)$, the semiconductor is called *non degenerate* and the balance relation (18.24) reduces to

$$N_C \exp\left(-\frac{E_C - E_F}{k_B T}\right) \simeq N_V \exp\left(-\frac{E_F - E_V}{k_B T}\right) + N_D. \tag{18.28}$$

From the exponential expressions of n and p it follows, using (18.13),

$$n\, p \simeq N_C\, N_V \exp\left(-\frac{E_C - E_V}{k_B T}\right) = N_C\, N_V \exp\left(-\frac{E_G}{k_B T}\right) = n_i^2. \tag{18.29}$$

Letting $N_D^+ = N_D$ in (18.23) and multiplying both sides of it by n provides, thanks to (18.29), an easy algebraic derivation of the electron and hole concentrations in the non-degenerate and spatially-uniform case:

$$n^2 - N_D\, n - n_i^2 = 0, \qquad n = \frac{N_D}{2} + \sqrt{\frac{N_D^2}{4} + n_i^2}, \qquad p = \frac{n_i^2}{n}. \tag{18.30}$$

The range of validity of (18.30) is quite vast; in silicon at room temperature the approximation (18.28) holds up to $N_D \simeq 10^{17}$ cm^{-3}; also, even for the lowest doping concentrations (about 10^{13} cm^{-3}) it is $N_D \gg n_i$, so that (18.30) can further be approximated as $n \simeq N_D$, $p \simeq n_i^2/N_D$. Thus, the concentration of majority carriers turns out to be independent of temperature, while that of minority carriers still depends on temperature through n_i. Assuming by way of example $N_D = 10^{15}$ cm^{-3} and taking a temperature such that $n_i = 10^{10}$ cm^{-3} (compare with Table 18.2), one finds $n \simeq 10^{15}$ cm^{-3}, $p \simeq 10^5$ cm^{-3}. This result is very important because it demonstrates that the concentration of minority carriers is negligible; as anticipated above, the inclusion of a suitable concentration of dopants makes the material's conductivity independent of temperature. The Arrhenius plot of $n(T)$ for an n-type semiconductor is shown in Fig. 18.9 in arbitrary units. The left part of the curve, called *intrinsic range*, corresponds to the situation where the intrinsic concentration prevails due to the high temperature; the plateau, called *saturation range*, corresponds to the situation where $n \simeq N_D$; finally, the right part of the curve, called *freeze-out range*, corresponds to the case where $n \simeq N_D^+ < N_D$, with N_D^+ decreasing as temperature decreases. From the practical standpoint, the important outcome is that the saturation region covers the range of temperatures within which the integrated circuits operate.[14]

[14] For instance, for silicon with $N_D = 10^{15}$ cm^{-3} the saturation region ranges from $T_{\min} \simeq 125$ K to $T_{\max} \simeq 370$ K [103, Sect. 2.4].

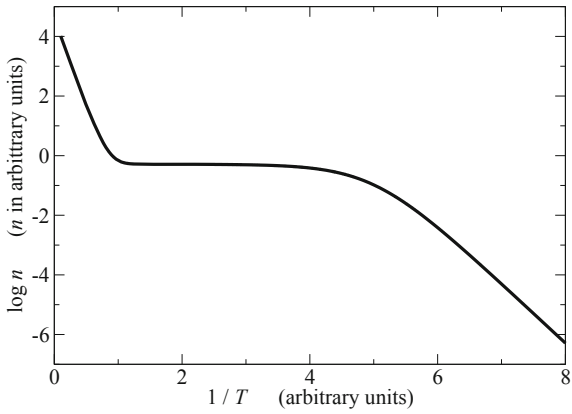

Fig. 18.9 Arrhenius plot of $n(T)$ for an n-type semiconductor, in arbitrary units

Coupling $n = N_D$ with the expression of n given by the left hand side of (18.28), and considering the values of N_C given in Table 18.2, yields

$$E_F = E_C - k_B\, T\, \log\left(\frac{N_C}{N_D}\right) < E_C, \tag{18.31}$$

whence, as anticipated,

$$\frac{\mathrm{d}E_F}{\mathrm{d}N_D} = \frac{k_B\, T}{N_D} > 0, \qquad \frac{\mathrm{d}E_F}{\mathrm{d}T} = k_B\, \log\left(\frac{N_D}{N_C}\right) = -\frac{E_C - E_F}{T} < 0. \tag{18.32}$$

From definition (18.26) of the Fermi potential it follows $E_C - E_F = E_C - E_{Fi} + q\,\varphi_F$; replacing the latter in $n = N_C\, \exp\left[-(E_C - E_F)/(k_B\, T)\right]$ yields an alternative expression of the equilibrium concentration and of the Fermi potential itself,

$$n = n_i\, \exp\left(-\frac{q\,\varphi_F}{k_B\, T}\right), \qquad \varphi_F = -\frac{k_B\, T}{q}\, \log\left(\frac{N_D}{n_i}\right) < 0. \tag{18.33}$$

18.4.2 Acceptor-Type Impurities

The analysis carried out in Sect. 18.4.1 is repeated here in a shorter form with reference to a substitutional impurity of the acceptor type like, e.g., boron (Fig. 18.10).

As the dopant concentration is small with respect to the semiconductor's, each boron atom is surrounded by silicon atoms. Boron has three electrons in the external shell: while these electrons form covalent bonds with silicon, the remaining unsaturated bond deforms the shape of the potential energy in the vicinity of the boron atom. This attracts an electron from a shared pair of a neighboring silicon. In other terms, to form four covalent bonds with silicon, boron generates an electron–hole pair as shown in the figure. In the $T \to 0$ limit, the holes are permanently bound to

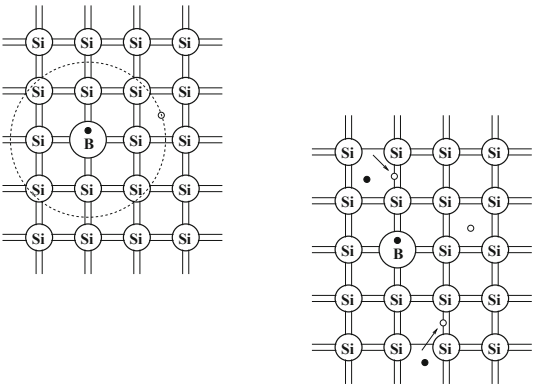

Fig. 18.10 Two-dimensional representation of the p-doped silicon lattice. The *upper-left* part of the figure shows the $T \to 0$ limit

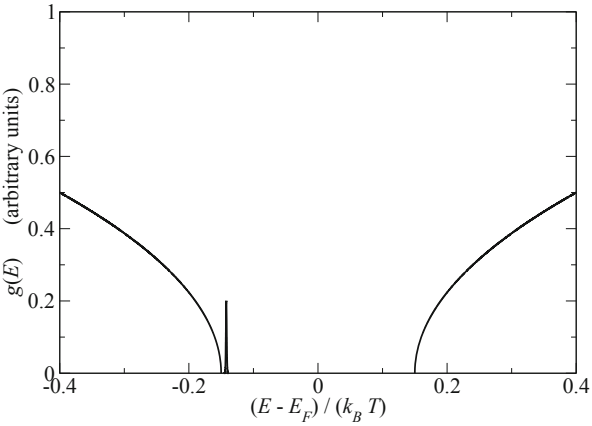

Fig. 18.11 Density of states in a p-doped semiconductor. The gap's extension is arbitrary and does not refer to any specific material

the atoms as in the intrinsic case; this situation is drawn in the upper-left part of the figure, where the hole is represented, in a particle-like picture, as orbiting around the boron atom. The orbit's radius is relatively large because the binding force is weak. The black dot inside the boron atom indicates the negative charge that balances the orbiting hole. If, instead, the temperature is brought to some finite value $T > 0$ by transferring energy from an external reservoir to the semiconductor, a fraction of the orbiting holes break the bond and become free to move within the material. At the same time, electrons belonging to shared pairs may also break their bonds, like in intrinsic silicon. This situation is depicted in the lower-right part of Fig. 18.10, where the free holes are represented with white dots. The negative charge within the boron atom remains trapped within the atom itself: thus, the boron atoms provide free holes to the lattice, but no electrons.

The analysis is better specified below, starting from the consideration of the density of states in a p-type semiconductor, shown in Fig. 18.11.

18.4.2.1 Concentration of Ionized Impurities (Acceptor Type)

Atoms like that of boron trap electrons from the lattice; for this reason they are also called *acceptors*. The concentration of acceptors is indicated with N_A and, in this section, is assumed to be uniform in space. When acceptor atoms are present, the equilibrium concentrations of electrons and holes are different from those of the intrinsic semiconductor; their derivation is identical to that of the intrisic or n-type case and yields again (18.17), with N_C, N_V given by (18.4, 18.8), respectively, and ξ_e, ξ_h given by (18.18).

In addition to the electron of the conduction band and holes of the valence band, a third population of particles must be accounted for, namely, that associated to the dopants. Let $N_A^- \leq N_A$ be the spatially-constant concentration of the acceptors that have released the orbiting hole to the lattice.[15] Equivalently, N_A^- may be thought of as the concentration of electrons that have been captured by the acceptor atoms;

$$N_A^- = \int_{\Delta E_A} \gamma_A(E)\, P_A(E)\, \mathrm{d}E, \tag{18.34}$$

where $\gamma_A(E)$ is the combined density of states produced by the acceptor atoms, ΔE_A the energy range where $\gamma_A(E) \neq 0$, and $P_A(E)$ the occupation probability of such states. The form of γ_A depends on the concentration N_A; for low or moderately-high values of the acceptor concentration, γ_A has the form shown in Fig. 18.11. Such a density of states can be approximated by a Dirac delta centered at an energy E_A, called *acceptor level*, positioned within the gap at a small distance from the edge E_V of the conduction band:

$$\gamma_A(E) \simeq N_A\, \delta(E - E_A). \tag{18.35}$$

The coefficient in (18.35) is such that (18.34) yields $N_A^- = N_A$ when $P_A = 1$. The distance of the acceptor level from the maximum of the valence band is calculated in Sect. 18.7.3; for the typical p-type dopants used with silicon it is $E_A - E_V \simeq 0.05$ eV for the heavy holes and $E_A - E_V \simeq 0.016$ eV for the light holes. Another important feature of the acceptor levels (still in the case of low or moderately-high values of the impurity concentration) is that they are localized in space, as schematically illustrated in Fig. 18.12. For this reason, the Fermi-Dirac statistics describing the equilibrium distribution of the electrons within the acceptor states is slightly different from that used for the band electrons, and reads

$$P_A(E) = \frac{1}{(1/d_A)\, \exp\left[(E - E_F)/(k_B\, T)\right] + 1}, \tag{18.36}$$

[15] A boron atom that has not released the orbiting hole is electrically neutral; when the orbiting hole breaks the bond and becomes free to move within the lattice, the atom becomes negatively ionized, because the release of a hole is actually the capture of a valence-band electron by the boron atom. The symbol N_A^- reminds one of that. The release of a second hole does not occur.

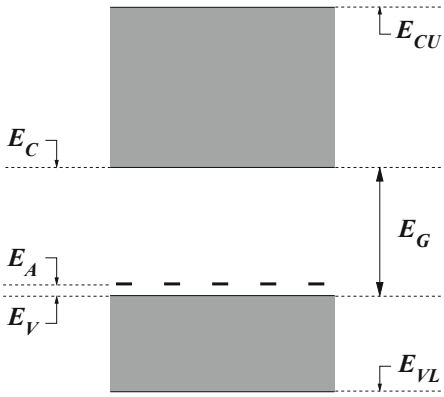

Fig. 18.12 Schematic representation of the acceptor states

with d_A the *acceptors' degeneracy coefficient*. In silicon, germanium, and gallium arsenide it is $d_A = 4$ [103, Sect. 2.4]. Combining (18.34) with (18.35) and (18.36) yields

$$N_A^- = \frac{N_A}{(1/d_A)\,\exp\left[(E_D - E_F)/(k_B\,T)\right] + 1}. \tag{18.37}$$

The position of the Fermi level is calculated by the same token as for the n-type dopant and is based on the balance relation

$$n + N_A^- = p, \tag{18.38}$$

whose limit for $N_A \to 0$ yields (18.11) as should be. Now, inserting (18.17, 18.37) into (18.38) after letting $\xi_A = (E_V - E_A)/(k_B\,T) < 0$, provides

$$N_C\,\Phi_{1/2}(\xi_e) + \frac{N_A}{(1/d_A)\,\exp\left(\xi_h - \xi_A\right) + 1} = N_V\,\Phi_{1/2}(\xi_h). \tag{18.39}$$

The latter, along with the relation $\xi_e + \xi_h = -E_G/(k_B\,T)$, form a system in the two unknowns ξ_e, ξ_h, whose solution determines the position of E_F with respect to the band edges E_C and E_V at a given temperature $T > 0$ and acceptor concentration N_A. It is easily found that, for a fixed temperature, the argument $\xi_h = (E_V - E_F)/(k_B\,T)$ increases when N_A increases. The demonstration is identical to that leading to (18.25). In the intrinsic case $N_A = 0$ it is $E_F = E_{Fi}$ and $\xi_h < 0$; as N_A increases, the Fermi level moves towards E_V, this making ξ_h less and less negative. At extremely high concentrations of the acceptor atoms, the Fermi level may reach E_V and even enter the valence band, this making ξ_h positive. As n, p, and N_A^- are non negative, it is intuitive that an increasing concentration of acceptor atoms makes p larger than n; for this reason, in a p-doped semiconductor holes are called *majority carriers* while electrons are called *minority carriers*.

To discuss from a qualitative standpoint the dependence of the Fermi level with respect to variations in temperature one may observe that, like in an n-doped material,

at very high temperatures the concentration of electron–hole pairs generated by the semiconductor prevails over the concentration of holes provided by the dopant atoms, so that the position of the Fermi level must be close to the intrinsic one; in contrast, when $T \to 0$ the Fermi level is positioned between E_V and E_A. In conclusion, for a p-doped semiconductor one expects that $\mathrm{d}E_F/\mathrm{d}T > 0$. Also, in a p-type semiconductor it is $E_F < E_{Fi}$; thus, the Fermi potential (18.26) is positive.

18.4.2.2 Complete Ionization and Non-Degenerate Condition (Acceptor Type)

When the Fermi level's position and temperature are such that

$$\xi_h - \xi_A < -1, \qquad E_A < E_F - k_B\,T, \tag{18.40}$$

the exponential term at the right hand side of (18.37) is negligible with respect to unity, so that $N_A^- \simeq N_A$. This condition is indicated with *complete ionization*. Note that, since $\xi_A < 0$, the complete-ionization condition in a p-doped semiconductor also implies $\xi_h < -1$; as a consequence, the approximation $\Phi_{1/2}(\xi_h) \simeq \exp(\xi_h)$ for the Fermi integral holds (Sect. C.13). Moreover, it is $\xi_e < \xi_h < -1$ because the Fermi level belongs to the lower half of the gap; thus, the same approximation applies to $\Phi_{1/2}(\xi_e)$ as well. In conclusion, the semiconductor is non degenerate and the balance relation (18.39) reduces to

$$N_C\,\exp\left(-\frac{E_C - E_F}{k_B\,T}\right) + N_A \simeq N_V\,\exp\left(-\frac{E_F - E_V}{k_B\,T}\right). \tag{18.41}$$

From the exponential expressions of n and p it follows, like in the n-type case, that the product of the equilibrium concentrations fulfills (18.29). The algebraic derivation of the electron and hole concentrations in the non-degenerate and spatially-uniform case is also similar:

$$p^2 - N_A\,p - n_i^2 = 0, \qquad p = \frac{N_A}{2} + \sqrt{\frac{N_A^2}{4} + n_i^2}, \qquad n = \frac{n_i^2}{p}, \tag{18.42}$$

with the further approximation $p \simeq N_A$, $n \simeq n_i^2/N_A$ holding within the low and moderately-high range of dopant concentrations, and around room temperature.

Coupling $p = N_A$ with the expression of p given by the right hand side of (18.41), and considering the values of N_V given in Table 18.2, yields

$$E_F = E_V + k_B\,T\,\log\left(\frac{N_V}{N_A}\right) > E_V, \tag{18.43}$$

whence, as anticipated,

$$\frac{\mathrm{d}E_F}{\mathrm{d}N_A} = -\frac{k_B\,T}{N_A} < 0, \qquad \frac{\mathrm{d}E_F}{\mathrm{d}T} = k_B\,\log\left(\frac{N_V}{N_A}\right) = \frac{E_F - E_V}{T} > 0. \tag{18.44}$$

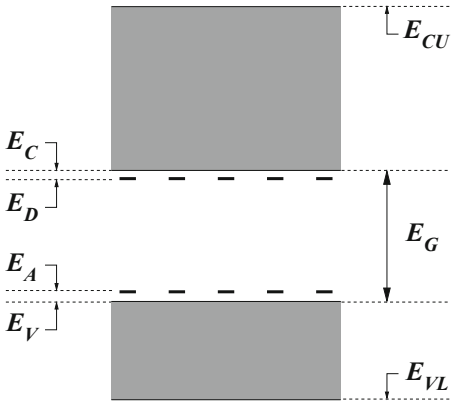

Fig. 18.13 Schematic representation of a semiconductor with both donor and acceptor states

From definition (18.26) of the Fermi potential it follows $E_F - E_V = E_{Fi} - E_V - q\,\varphi_F$; replacing the latter in $p = N_V\,\exp\left[-(E_F - E_V)/(k_B\,T)\right]$ yields an alternative expression of the equilibrium concentration and of the Fermi potential itself,

$$p = n_i\,\exp\left(\frac{q\,\varphi_F}{k_B\,T}\right), \qquad \varphi_F = -\frac{k_B\,T}{q}\,\log\left(\frac{N_A}{n_i}\right) > 0. \tag{18.45}$$

18.4.3 Compensation Effect

The architecture of semiconductor devices is such that donor and acceptor dopants are present in the same region (Fig. 18.13). Letting N_D, N_A be the corresponding concentrations, still assumed uniform in space, the equilibrium concentrations of electrons and holes are expressed by (18.17) as in the other cases discussed above, with N_C, N_V and ξ_e, ξ_h given by (18.4, 18.8) and (18.18), respectively. In turn, the concentrations of ionized donors and acceptors are given by (18.22) and (18.37), respectively. The balance equation reads

$$n + N_A^- = p + N_D^+, \tag{18.46}$$

namely, $N_C\,\Phi_{1/2}(\xi_e) + N_A\,P_A(E_A) = N_V\,\Phi_{1/2}(\xi_h) + N_D\,[1 - P_D(E_D)]$. One observes that if $N_D^+ = N_A^-$, the balance equation (18.46) coincides with that of an intrinsic semiconductor (compare with (18.11)). As a consequence, the position of the Fermi level is given by (18.16), and the electrons released by the donor atoms are trapped by the acceptor ones. In this case the semiconductor is *fully compensated*. If, instead, it is $N_D^+ \neq N_A^-$, the semiconductor is *partially compensated*, and one must distinguish between two cases of the balance relation (18.46), depending on the sign of the *net ionized impurity concentration* $N = N_D^+ - N_A^-$. In the first case,

$$N > 0, \qquad n = p + N, \tag{18.47}$$

the balance relation is identical to that of an n-doped semiconductor (compare with (18.23)), with an effective donor dopant equal to N. In the second case,

$$N < 0, \qquad n + |N| = p, \tag{18.48}$$

the balance relation is identical to that of a p-doped semiconductor (compare with (18.38)), with an effective acceptor dopant equal to $|N|$. In the non-degenerate case, (18.29) still holds. If complete ionization also occurs, then $N = N_D - N_A$; when the donor-type dopant prevails, electrons are the majority carriers, and the same calculation as that leading to (18.30) yields

$$n = \frac{N}{2} + \sqrt{\frac{N^2}{4} + n_i^2}. \tag{18.49}$$

If $N \gg n_i$, then

$$n \simeq N, \qquad p \simeq \frac{n_i^2}{N}, \qquad \varphi_F = -\frac{k_B\, T}{q} \log\left(\frac{N}{n_i}\right) < 0. \tag{18.50}$$

If, on the contrary, the acceptor-type dopant prevails, holes are the majority carriers, and the same calculation as that leading to (18.42) yields

$$p = -\frac{N}{2} + \sqrt{\frac{N^2}{4} + n_i^2}. \tag{18.51}$$

If $-N \gg n_i$, then

$$p \simeq -N = |N|, \qquad n \simeq \frac{n_i^2}{|N|}, \qquad \varphi_F = \frac{k_B\, T}{q} \log\left(\frac{|N|}{n_i}\right) < 0. \tag{18.52}$$

18.5 Non-Uniform Distribution of Dopants

This section deals with the more realistic case where the donor concentration N_D, the acceptor concentration N_A, or both, depend on position. A qualitative reasoning shows that the balance equations (18.23), (18.38), or (18.46) do not hold any more. To show this, consider the simple case where only the n-type dopant is present, $N_D = N_D(\mathbf{r})$, $N_A = 0$, and select two nearby positions, say, $\mathbf{r}_1$ and $\mathbf{r}_2$, in the limiting case $T \to 0$; then, let the temperature increase such that a fraction of the donor atoms ionizes. If $N_D(\mathbf{r}_1) \neq N_D(\mathbf{r}_2)$, the numbers of electrons transiting to the conduction band at $\mathbf{r}_1$ and $\mathbf{r}_2$ is different, so that the concentration in one position, say, $\mathbf{r}_1$ is larger than that in the other; thus, some electrons diffuse[16] from the former position to the latter. On the other hand, as the position of the positive charges

[16] Diffusive transport is introduced in Sect. 23.3.

within the ionized donors is fixed, the spatial rearrangement of the conduction-band electrons unbalances the negative charges with respect to the positive ones, so that the local charge density differs from zero. The same reasoning applies when both donor and acceptor dopants are present, so that in a general non-uniform case it is

$$\varrho(\mathbf{r}) = q\,\left[p(\mathbf{r}) - n(\mathbf{r}) + N_D^+(\mathbf{r}) - N_A^-(\mathbf{r})\right] \neq 0. \tag{18.53}$$

A non-vanishing charge density produces in turn an electric field $\mathbf{E} = \mathbf{E}(\mathbf{r})$ whose action balances the diffusion. In conclusion, the equilibrium condition is kept by an exact balance of the two transport mechanisms. Considering that the equilibrium condition is time-independent, the Maxwell equations in the semiconductor reduce to $\varepsilon_{sc}\,\mathrm{div}\mathbf{E} = \varrho$, with $\mathbf{E} = -\mathrm{grad}\varphi$ (Sect. 4.4), φ and ε_{sc} being the electric potential and semiconductor permittivity, respectively.[17] In other terms, the electric field due to the non-uniformity of ϱ is found by solving the Poisson equation.

The effect onto the total energy of the electrons due to the presence of an electric field also influences the statistical distribution of the electrons in the energy states. It will be demonstrated in Sect. 19.2.2 that the statistical distribution to be used here is a modified form of the Fermi-Dirac statistics where E is replaced with $E - q\,\varphi(\mathbf{r})$:

$$P(E,\mathbf{r}) = \frac{1}{\exp\left[(E - q\,\varphi(\mathbf{r}) - E_F)\,/\,(k_B\,T)\right] + 1}; \tag{18.54}$$

similarly, (18.21) and (18.36) become

$$P_{D(A)}(E,\mathbf{r}) = \frac{1}{(1/d_{D(A)})\,\exp\left[(E - q\,\varphi(\mathbf{r}) - E_F)/(k_B\,T)\right] + 1}. \tag{18.55}$$

Note that the calculations leading to the concentrations are carried out in the same manner as in Sects. 18.4.1, 18.4.2, because they involve integrals over energy only. As a consequence, (18.17) generalize to

$$n(\mathbf{r}) = N_C\,\Phi_{1/2}\left(\xi_e(\mathbf{r})\right), \qquad \xi_e(\mathbf{r}) = -\frac{\zeta_e(\mathbf{r})}{k_B\,T} = -\frac{E_C - E_F - q\,\varphi(\mathbf{r})}{k_B\,T}, \tag{18.56}$$

and

$$p(\mathbf{r}) = N_V\,\Phi_{1/2}\left(\xi_h(\mathbf{r})\right), \qquad \xi_h(\mathbf{r}) = -\frac{\zeta_h(\mathbf{r})}{k_B\,T} = -\frac{E_F + q\,\varphi(\mathbf{r}) - E_V}{k_B\,T}. \tag{18.57}$$

Similarly, (18.22, 18.34) become

$$N_D^+(\mathbf{r}) = N_D(\mathbf{r})\,\left[1 - P_D(E_D,\mathbf{r})\right], \qquad N_A^-(\mathbf{r}) = N_A(\mathbf{r})\,P_A(E_A,\mathbf{r}). \tag{18.58}$$

Note that the summands at the right hand side of (18.53) depend on position through the electric potential φ and also through the explicit dependence of N_D and N_A; as a

[17] Note that the material's permittivity must be used here instead of vacuum's. This is coherent with the use of charge density and, in a non-equilibrium situation, of current density, which entail averages over volumes of space.

consequence, inserting the expression of the charge density into the right hand side of the Poisson equation yields

$$-\varepsilon_{sc}\,\nabla^2\varphi = \varrho(\varphi, \mathbf{r}), \tag{18.59}$$

namely, a second-order, partial differential equation in the unknown φ, with position-dependent coefficients. The equation must be supplemented with suitable boundary conditions.[18] Equation (18.59) is the generalization of the balance equation (18.46) to the non-uniform case: in the case of (18.46), the problem is algebraic and yields E_F, whereas in the case of (18.59) it is differential and yields $E_F + q\,\varphi$. After solving (18.59) one reconstructs n, p, N_D^+, and N_A^- at each point through (18.56), (18.57), and (18.58). In the non-degenerate case, (18.56) becomes

$$n = n^{(0)}\,\exp\left(\frac{q\,\varphi}{k_B\,T}\right), \quad n^{(0)} = N_C\,\exp\left(\frac{E_F - E_C}{k_B\,T}\right) = n_i\,\exp\left(\frac{-q\varphi_F}{k_B\,T}\right), \tag{18.60}$$

with $n^{(0)}$ the value of the electron concentration in the position(s) where $\varphi = 0$. The last expression of $n^{(0)}$ is obtained by combining definition (18.26) with first relation in (18.12). Similarly,

$$p = p^{(0)}\,\exp\left(\frac{-q\,\varphi}{k_B\,T}\right), \quad p^{(0)} = N_V\,\exp\left(\frac{E_V - E_F}{k_B\,T}\right) = n_i\,\exp\left(\frac{q\varphi_F}{k_B\,T}\right). \tag{18.61}$$

From (18.60, 18.61) one finally obtains

$$n = n_i\,\exp\left[\frac{q\,(\varphi - \varphi_F)}{k_B\,T}\right], \qquad p = n_i\,\exp\left[\frac{q\,(\varphi_F - \varphi)}{k_B\,T}\right]. \tag{18.62}$$

One observes that (18.29) still holds, namely, $n\,p = n^{(0)}\,p^{(0)} = n_i^2$: in the non-degenerate case, the equilibrium product does not depend on position. If complete ionization also occurs, then $N_D^+(\mathbf{r}) = N_D(\mathbf{r})$, $N_A^-(\mathbf{r}) = N_A(\mathbf{r})$: the ionized-dopant concentrations do not depend on the electric potential.

18.6 Band-Gap Narrowing

When the dopant concentration is large, the density of states associated to the dopant atoms can no longer be described as in Figs. 18.7, 18.11. Rather, considering by way of example an n-type dopant, the form of the density of states is similar to that shown in Fig. 18.14, namely, it overlaps the lower portion of the conduction band forming

[18] Such boundary conditions must be coeherent with the choice of the zero point of the total energy. An example is given in Sect. 21.2.

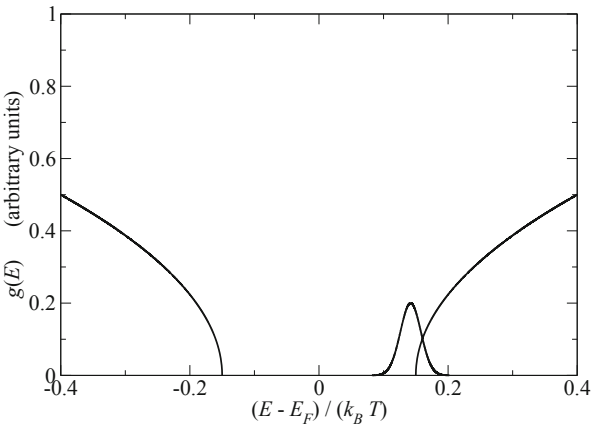

Fig. 18.14 Density of states in an n-doped semiconductor, where the high concentration of the dopant produces the band-gap narrowing. The gap's extension is arbitrary and does not refer to any specific material

the so-called *impurity band*. In addition, the dopant atoms are close to each other, to the extent that the probability of tunneling of an electron, from a neutral to a nearby, ionized donor atom, is not negligible (Sect. 18.7.2). In a non-equilibrium condition, the tunneling electrons give rise to a current density; the phenomenon is referred to as *impurity-band conduction*. From the practical standpoint, the union of the conduction and impurity bands is viewed as a broader conduction band whose lower edge is shifted with respect to the undoped, or moderately-doped case; the effect is also called *band-gap narrowing*. The analysis is similar when a large concentration of acceptor atoms is present. In conclusion, band-gap narrowing is in general produced by the lowering of the conduction-band edge, $\Delta E_C(\mathbf{r}) > 0$, combined with the lifting of the valence-band edge, $\Delta E_V(\mathbf{r}) > 0$. Both quantities are position dependent because in general the dopant concentrations are such.

Indicating with E_{Ci}, E_{Vi} the lower edge of the conduction band and, respectively, the upper edge of the valence band in the undoped or moderately-doped case, and observing that the variations due to heavy doping are positive, one has for the actual positions of the band edges:

$$E_C(\mathbf{r}) = E_{Ci} - \Delta E_C(\mathbf{r}), \qquad E_V(\mathbf{r}) = E_{Vi} + \Delta E_V(\mathbf{r}), \tag{18.63}$$

whence

$$E_G(\mathbf{r}) = E_{Gi} - \Delta E_G(\mathbf{r}), \quad E_{Gi} = E_{Ci} - E_{Vi}, \quad \Delta E_G = \Delta E_C + \Delta E_V > 0. \tag{18.64}$$

To calculate the carrier concentrations when band-gap narrowing is present, one must replace E_C with $E_{Ci} - \Delta E_C(\mathbf{r})$ in the second relation of (18.56), and E_V with $E_{Vi} + \Delta E_V(\mathbf{r})$ in the second relation of (18.57), to find

$$\xi_e(\mathbf{r}) = -\frac{E_{Ci} - E_F - q\,\varphi(\mathbf{r})}{k_B\,T} + \frac{\Delta E_C(\mathbf{r})}{k_B\,T}, \tag{18.65}$$

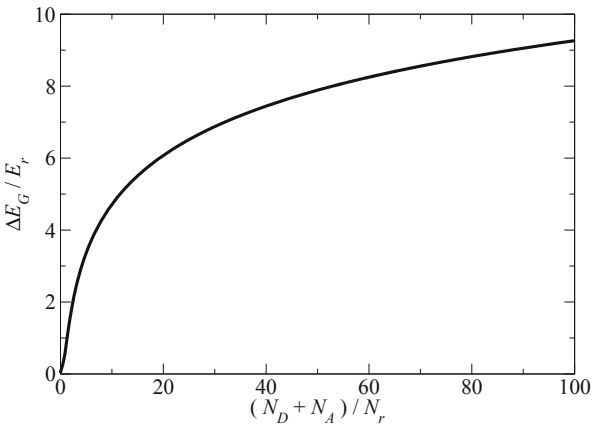

Fig. 18.15 Band-gap narrowing as a function of the total doping concentration, in normalized form, using the experimental expression 18.67

$$\xi_h(\mathbf{r}) = -\frac{E_F - E_{Vi} + q\,\varphi(\mathbf{r})}{k_B\,T} + \frac{\Delta E_V(\mathbf{r})}{k_B\,T}. \tag{18.66}$$

Band-gap narrowing makes ξ_e, ξ_h to increase with respect to the moderately-doped case; as a consequence, the equilibrium carrier concentrations n and p are larger as well. As mentioned in Sect. 18.3, band gap is measured by either electrical or optical methods. The results of gap's measurements, that show that E_G decreases when the dopant concentration exceeds some limiting value N_r, are usually rendered in compact form by means of interpolating expressions, an example of which is [100, 101]

$$\Delta E_G = E_r \left(F + \sqrt{F^2 + 0.5}\right), \qquad F = \log\left(\frac{N_D + N_A}{N_r}\right), \tag{18.67}$$

with $E_r = 9$ meV, $N_r = 10^{17}$ cm^{-3}. The function described by (18.67) is shown in normalized form in Fig. 18.15. Expressions like (18.67) describe the cumulative effect of the total doping concentration, without distinguishing between the donor or acceptor contribution to band-gap narrowing. For this reason, when band-gap narrowing is accounted for in numerical calculations, ΔE_G is equally distributed between the two bands, namely, $\Delta E_C = \Delta E_V = \Delta E_G/2$.

It is interesting to note that the onset of band-gap narrowing corresponds to a total dopant concentration of about 10^{17} cm^{-3}, where the non-degeneracy condition still holds. As a consequence, a range of dopant concentrations exists where the exponential approximation can be used for the equilibrium carrier concentrations, whereas the band-gap narrowing effect must be accounted for.[19] The non-degeneracy condition reads in this case

$$E_{Ci} - E_F - q\,\varphi - \Delta E_C > k_B\,T, \qquad E_F - E_{Vi} + q\,\varphi - \Delta E_V > k_B\,T. \tag{18.68}$$

[19] Such a range may be quite large if one considers the compensation effect (Sect. 18.4.3).

Remembering (18.62), the equilibrium concentrations become

$$n = n_e \exp\left[\frac{q\ (\varphi - \varphi_F)}{k_B\, T}\right], \qquad n_e = n_i \exp\left(\frac{\Delta E_C}{k_B\, T}\right) > n_i, \tag{18.69}$$

$$p = p_e \exp\left[\frac{q\ (\varphi_F - \varphi)}{k_B\, T}\right], \qquad p_e = n_i \exp\left(\frac{\Delta E_V}{k_B\, T}\right) > n_i, \tag{18.70}$$

where $p_e = n_e$ on account of $\Delta E_V = \Delta E_C$. The common value n_e is called *effective intrinsic concentration*. The equilibrium product then reads

$$n\, p = n_e^2, \qquad n_e^2 = n_i^2 \exp\left(\frac{\Delta E_G}{k_B\, T}\right). \tag{18.71}$$

18.7 Complements

18.7.1 Si, Ge, GaAs in the Manufacturing of Integrated Circuits

As noted in Sect. 18.3, silicon, germanium, and gallium arsenide have similar effective densities of states, but different gap extensions (Table 18.1); in this respect, silicon is considered as a reference material, so that germanium is indicated as a *narrow-gap material* while gallium arsenide is indicated as a *wide-gap material*. The differences in the gap extension produce huge differences in the intrinsic concentration n_i (Table 18.2); the latter, in turn, has a strong influence on the functioning of the integrated circuits. In fact, the saturation current of a *p-n* junction is proportional to n_i^2 (Sect. 21.3.1); as a consequence, this parameter determines the current of the junction when a reverse bias is applied to it. When many junctions are present, as is typically the case in integrated circuits, the inverse currents may build up and give rise to a substantial parasitic current. From this standpoint, gallium arsenide is preferable with respect to silicon, which in turn is preferable with respect to germanium. Gallium arsenide is also preferable because of the smaller effective mass of the electrons, which makes the electron mobility larger (Sect. 17.6.6). On the other hand, silicon is much less expensive; in fact it is the second most abundant element in Earth's crust (the first one is oxygen); gallium, germanium, and arsenic are much rarer.

The historical development of the semiconductor-device manufacture has followed, instead, a different path. Until the mid sixties of the last century, germanium was preferred; the reason for this was that a technological process, able to purify the material to the level required by the electronic industry, was available for germanium first. As soon as the purification method became available for silicon as well, the latter replaced germanium in the fabrication of semiconductor devices and, soon after, of integrated circuits. The silicon technology developed with a steady pace, giving rise to decades of exponential miniaturization and development in the integrated-circuit manufacture. The miniaturization of gallium-arsenide-based circuits did not

proceed with the same pace because the technology of compound materials is more complicate.

In 1980, practically 100 % of the worldwide market share of integrated circuit was silicon based, almost equally distributed between the bipolar and MOSFET technologies [13]. In the following years the MOSFET technology became dominant, reaching a market share of 88 % in the year 2000; of the remaining 12 %, the bipolar, silicon-based technology kept an 8 % share, while the remaining 4 % was taken by integrated circuits using III-V compounds.

In 1989, germanium was introduced again in the silicon integrated-circuit technology to form silicon-germanium alloys ($Si_{1-x}Ge_x$). The alloy makes a flexible band-gap tuning possible; it is used for manufacturing heterojunction bipolar transistors, yielding higher forward gain and lower reverse gain than traditional bipolar transistors. Another application of the alloy is in the Silicon-Germanium-On-Insulator (SGOI) technology. The difference in the lattice constants of germanium and silicon induces a strain in the material under the gate, that makes electron mobility to increase.

18.7.2 Qualitative Analysis of the Impurity Levels

A qualitative analysis of the impurity levels may be carried out basing on a modified version of the Kronig–Penney model discussed in Sect. 17.8.4. To this purpose, consider the case of donor atoms placed at equal distances in a one-dimensional lattice, like that of Fig. 18.16. The deeper wells are those introduced by the dopants, while the finer structure above the x axis is due to the semiconductor nuclei. Note that the relative distances in the figure are not realistic and are used only for illustrative purposes; assuming in fact that the structure represented a cross section of a three-dimensional semiconductor, where a uniform donor concentration N_D is present, the distance between two neighboring impurity atoms is found to be $(1/N_D)^{1/3}$. If the semiconductor's concentration is N_{sc}, the ratio $(N_{sc}/N_D)^{1/3}$ indicates how many semiconductor atoms are present in the interval between two neighboring impurities. Considering silicon by way of example ($N_{sc} = 5 \times 10^{22}$ cm^{-3}), and taking $N_D = 5 \times 10^{16}$ cm^{-3} yields $(N_{sc}/N_D)^{1/3} = 100$.

With this provision, let c be the width of the barrier separating two dopant-induced wells, and consider a negative value E' of the electron's energy. If c is large, the probability that the electron tunnels from a well to an adjacent, empty one is negligibly small; in this case, the electron is confined within the well where the localization probability is the largest, and the energy states $E' < 0$ are similar to those of a single well (the energy states of the finite rectangular well are worked out in Sect. 11.5). The lowest state of the well is the ground state E_D shown in Fig. 18.16. If, instead, a positive energy state E is considered, the wave function is extended over the whole lattice like in the Kronig–Penney model. In summary, the addition of dopant atoms with a low or moderately high concentration, such that their mutual distance c is large, provides a set of energy states that adds up to the band structure of the intrinsic

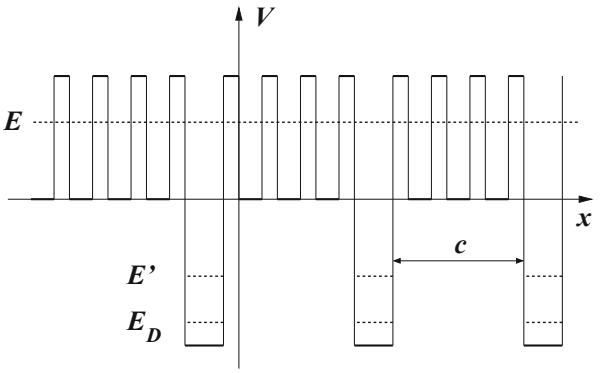

Fig. 18.16 Potential energy in the Kronig–Penney model modified to account for impurity atoms

semiconductor. The states introduced by the dopants are localized in space at the positions of the dopant atoms, and are distributed in energy as discrete levels whose mutual distances depend on the form of the well.

In turn, the states with positive energy have a structure similar to that of the intrinsic semiconductor, because in this respect the dopant atoms have little effect on the lattice; in Fig. 18.16, the lower edge E_C of the conduction band coincide with $E = 0$. As said above, electrons belonging to the dopant atoms can not move as long as they are confined within wells, because tunneling is precluded; on the other hand, they may be promoted to conduction-band states by absorbing energy in a collision, and become band electrons, that is, mobile. Conversely, a band electron may loose energy due to a collision, and be trapped in an empty well.

When, due to an increasing dopant concentration, the width c of the barrier becomes smaller, the transmission coefficient increases and the electrons belonging to the wells have a non-negligible probability of moving from a well to another. When the structure of the wells becomes finer, the description of their energy states becomes more similar to that applicable to the intrinsic semiconductor: for an infinite structure one obtains a continuum of states, that fill up the well and connect to those of the conduction band. This explains the band-gap narrowing phenomenon introduced in Sect. 18.6.

18.7.3 Position of the Impurity Levels

To determine the position of the impurity levels for the case of a low or moderately high impurity concentration of the donor type, one considers the dispersion relation $E(\mathbf{k})$ of the conduction band of the intrinsic semiconductor, in the parabolic-band approximation (17.57, 17.58):

$$E(\mathbf{k}) \simeq E_C + \sum_{i=1}^{3} \frac{\hbar^2}{2\,m_{ia}} (k_i - k_{ia})^2, \qquad \frac{1}{m_{ia}} = \frac{1}{\hbar^2} \left(\frac{\partial^2 E}{\partial k_i^2} \right)_a > 0. \qquad (18.72)$$

Now, assume that a donor-type impurity is added in the origin, and that the impurity is ionized. As a consequence, it gives rise to a hydrogenic-like potential energy[20] of the form $V = -q^2/(4\pi\,\varepsilon_{\rm sc}\,r)$. The effect of V may be considered as a local perturbation over the band structure, so that when the ionized impurity is present the total energy of an electron becomes $H = E(\mathbf{k}) + V(r)$, with $E(\mathbf{k})$ the same as in the unperturbed case. Shifting the origin of $\mathbf{k}$ to $\mathbf{k}_a$, and replacing k_i with $-\mathrm{i}\,\partial/\partial x_i$, one obtains the Hamiltonian operator[21]

$$\mathcal{H} \simeq E_C - \sum_{i=1}^{3} \frac{\hbar^2}{2\,m_{ia}} \frac{\partial^2}{\partial x_i^2} - \frac{q^2}{4\pi\,\varepsilon_{\rm sc}\,r}. \tag{18.73}$$

To estimate the eigenvalues of $\mathcal{H}$ one may replace each m_{ia} with the average $m_{ea} = (m_{1a}\,m_{2a}\,m_{3a})^{1/3}$ (Sect. 17.6.3), to obtain

$$-\frac{\hbar^2}{2\,m_{ea}}\,\nabla^2 v_n - \frac{q^2}{4\pi\,\varepsilon_{\rm sc}\,r}\,v_n = (E'_n - E_C)\,v_n, \tag{18.74}$$

with v_n the eigenfunction. Apart from the coefficients, (18.74) is identical to the Schrödinger equation for the Coulomb case treated in Sect. 13.5.2, whose eigenvalues are given by (13.49). It follows

$$E'_n = E_C - \frac{m_{ea}}{2\,\hbar^2}\left(\frac{q^2}{4\pi\,\varepsilon_{\rm sc}}\right)^2 \frac{1}{n^2}, \qquad n = 1, 2, \ldots \tag{18.75}$$

Thus, the donor impurity provides infinite levels from the minimum E'_1 to the maximum E_C. Considering the case of silicon by way of example, one has $m_{1a} = m_l = 0.97\,m_0$, $m_{2a} = m_{3a} = m_t = 0.19\,m_0$ (Table 17.4), with $m_0 \simeq 9.11 \times 10^{-31}$ kg the rest mass of the electron, whence $m_{ea} = 0.33\,m_0$. Letting $E_D = E'_1$ and using $\varepsilon_{\rm sc} \simeq 11.7\,\varepsilon_0$, with $\varepsilon_0 = 8.854 \times 10^{-14}$ F cm^{-1} the vacuum permittivity, one finds the ionization energy of the donor impurity in silicon:

$$E_C - E_D = \frac{m_{ea}}{2\,\hbar^2}\left(\frac{q^2}{4\pi\,\varepsilon_{\rm sc}}\right)^2 \simeq 32.8 \quad \text{meV}. \tag{18.76}$$

The analysis is similar for an ionized, acceptor-type impurity. The hydrogenic-like potential energy becomes $V = q^2/(4\pi\,\varepsilon_{\rm sc}\,r)$, and the dispersion relation around a maximum $\mathbf{k}_a$ of the valence band reads

$$E(\mathbf{k}) \simeq E_V - \sum_{i=1}^{3} \frac{\hbar^2}{2m_{ia}}\,(k_i - k_{ia})^2, \qquad \frac{1}{m_{ia}} = -\frac{1}{\hbar^2}\left(\frac{\partial^2 E}{\partial k_i^2}\right)_a > 0. \tag{18.77}$$

[20] Compare with Sect. 17.6.6. Like in Sect. 18.5, the semiconductor permittivity is used instead of that of vacuum, because the wave function of an electron subject to the force due to V extends over several lattice cells.

[21] More comments about the procedure of obtaining an operator from a simplified form of the eigenvalues of a more general operator are made in Sect. 19.2.

When the impurity is present the Hamiltonian operator becomes

$$\mathcal{H} \simeq E_V + \sum_{i=1}^{3} \frac{\hbar^2}{2\,m_{ia}} \frac{\partial^2}{\partial x_i^2} + \frac{q^2}{4\,\pi\,\varepsilon_{\rm sc}\,r}. \tag{18.78}$$

Again, to estimate the eigenvalues one replaces each m_{ia} with the average m_{ha}. One finds

$$-\frac{\hbar^2}{2\,m_{ha}} \nabla^2 v_n - \frac{q^2}{4\,\pi\,\varepsilon_{\rm sc}\,r}\, v_n = (E_V - E_n'')\, v_n, \tag{18.79}$$

whence

$$E_n'' = E_V + \frac{m_{ha}}{2\,\hbar^2} \left(\frac{q^2}{4\,\pi\,\varepsilon_{\rm sc}}\right)^2 \frac{1}{n^2}, \tag{18.80}$$

$n = 1, 2, \ldots$. The acceptor impurity provides infinite levels from the minimum E_V to the maximum E_1''. Letting $E_A = E_1''$ one finds the ionization energy $E_A - E_V$ of the donor impurity. Taking again silicon by way of example, and using the values $m_{hh} = 0.5\,m_0$, $m_{hl} = 0.16\,m_0$ from Table 17.3, one finds $E_A - E_V = 49.7$ meV for the heavy holes and $E_A - E_V = 15.9$ meV for the light holes.

The ionization energies of phosphorus and boron *in vacuo* are about 10.5 eV and 8.3 eV, respectively, that is, much higher than those calculated here. The strong difference is ascribed to the presence of the silicon crystal: a comparison between (13.49) and (18.75) or (18.80) shows in fact that the coefficients of $1/n^2$ in the crystal case are much smaller than that *in vacuo*, due to the presence of the effective mass in the numerator and of the square of the material permittivity in the denominator. The small distance between the ground state of the impurity atoms and the edge of the band explains the ease with which the dopants ionize at room temperature.