Chapter 7
Electronic Defect States

Über Halbleiter sollte man nicht arbeiten, das ist eine Schweinerei, wer weiß ob es überhaupt Halbleiter gibt.
One should not work on semiconductors. They are a mess. Who knows whether semiconductors exist at all.
W. Pauli, 1931 [482]

Abstract After the carrier statistics for intrinsic conduction and general doping principles, donors and acceptors, compensation and high doping effects are treated in detail. The concept of quasi-Fermi levels is introduced. Finally for deep levels and their thermodynamics general remarks and several examples are given.

7.1 Introduction

One centimeter cube of a semiconductor contains about 5×10^{22} atoms. It is practically impossible to achieve perfect purity. Typical low concentrations of impurity atoms are in the 10^{12}–$10^{13}\,\mathrm{cm}^{-3}$ regime. Such a concentration corresponds to a purity of 10^{-10}, corresponding to about one alien in the world's human population. In the beginning of semiconductor research the semiconductors were so impure that the actual semiconducting properties could only be used inefficiently. Nowadays, thanks to large improvements in high-purity chemistry, the most common semiconductors, in particular silicon, can be made so pure that the residual impurity concentration plays no role in the physical properties. However, the most important technological step for semiconductors is *doping*, the controlled incorporation of impurities, in order to manage the semiconductor's conductivity. Typical impurity concentrations used in doping are 10^{15}–$10^{20}\,\mathrm{cm}^{-3}$. A milestone in the understanding of doping and the spreading of semiconductor technology was the 1950 textbook by Shockley [483].

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_7

7.2 Carrier Concentration

Generally, the density of electrons in the conduction band is given by

$$n = \int_{E_C}^{\infty} D_e(E)\, f_e(E)\, \mathrm{d}E, \tag{7.1}$$

and accordingly the density of holes in the valence band is

$$p = \int_{-\infty}^{E_V} D_h(E)\, f_h(E)\, \mathrm{d}E. \tag{7.2}$$

The energy of the top of the valence band is denoted by E_V, the bottom of the conduction band as E_C. The distribution function for holes is $f_h = 1 - f_e$. Thus,

$$f_h(E) = 1 - \frac{1}{\exp\left(\frac{E-E_F}{kT}\right)+1} = \frac{1}{\exp\left(-\frac{E-E_F}{kT}\right)+1}. \tag{7.3}$$

If several hole bands (hh, lh, so) are considered, the same distribution is valid for all hole bands in thermal equilibrium.

We assume parabolic band edges, i.e. effective masses m_e and m_h for electrons and holes, respectively. The density of states (per volume) in the conduction and valence bands is given by (6.70) and (6.71).

In thermodynamic equilibrium, the distribution function $f_e(E)$ for electrons is given by the Fermi–Dirac distribution (Fermi function) $f_e(E)$ (E.22)

$$f_e(E) = \frac{1}{\exp\left(\frac{E-E_F}{kT}\right)+1}, \tag{7.4}$$

If the Boltzmann distribution (E.23) is a good approximation, the carrier distribution is called *nondegenerate*. If the Fermi distribution needs to be invoked, the carrier ensemble is called *degenerate*. If the Fermi level is within the band, the ensemble is highly degenerate.

If the Boltzmann approximation (E.23) cannot be applied, i.e. at high temperatures or for very small band gaps, the integral over Df cannot be analytically evaluated. In this case the Fermi integral is needed that is defined[1] as

$$F_n(x) = \frac{2}{\sqrt{\pi}} \int_0^{\infty} \frac{y^n}{1+\exp(y-x)}\, \mathrm{d}y. \tag{7.5}$$

[1] Equation (7.5) is restricted to $n > -1$. A form without restriction is $\mathcal{F}_n(x) = \frac{1}{\Gamma(n+1)} \int_0^{\infty} \frac{y^n}{1+\exp(y-x)}\, \mathrm{d}y$. The factor $2/\sqrt{\pi}$ is often omitted but must be then added explicitly in, e.g. (7.6).

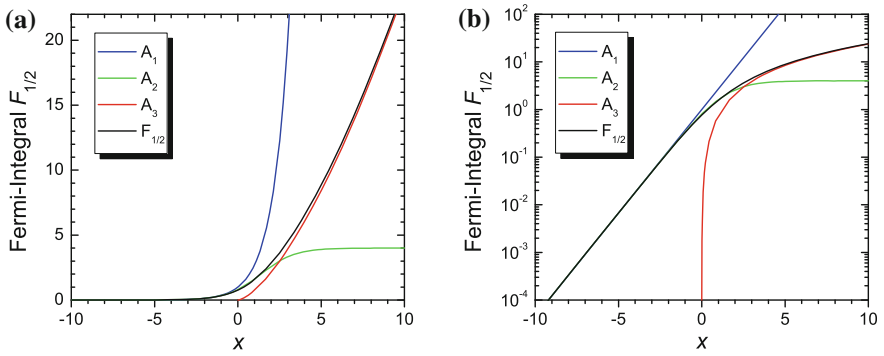

Fig. 7.1 Fermi integral $\hat{F}_{1/2} = (\sqrt{\pi}/2)F_{1/2}$ with approximations in three regions of the argument: $A_1(x) = (\sqrt{\pi}/2)\exp(x)$ for $x < 2$, $A_2(x) = (\sqrt{\pi}/2)(1/4 + \exp(-x))^{-1}$ for $-2 < x < 2$, $A_3(x) = 2/3x^{3/2}$ for $x > 2$. (**a**) linear, (**b**) semilogarithmic plot

In the present case of bulk materials $n = 1/2$. For large negative argument, i.e. $x < 0$ and $|x| \gg 1$, $F_{1/2}(x) \approx \exp(x)$, which is the Boltzmann approximation. $F_{1/2}(0) = 0.765\ldots \approx 3/4$. For large argument, i.e. $x \gg 1$, $F_{1/2}(x) \approx (2/\sqrt{\pi})(2/3)$ $x^{3/2}$. Such fairly simple approximations are plotted in Fig. 7.1 in comparison with the Fermi integral. For computations, analytical [484–486] or numerical approximations [487, 488] are used.

The derivative of the Fermi integral is given by $F_n'(x) = nF_{n-1}(x)$, $n > 0$. For $n = 0$, i.e. a two-dimensional system, the integral can be executed explicitly, $F_0(x) = (2/\sqrt{\pi})\ln\left[1 + \exp(x)\right]$.

With the Fermi integral $F_{1/2}$ (7.10) and (7.11) the free-carrier densities can be written as

$$n = N_\mathrm{C}\, F_{1/2}\left(\frac{E_\mathrm{F} - E_\mathrm{C}}{kT}\right) \tag{7.6}$$

$$p = N_\mathrm{V}\, F_{1/2}\left(-\frac{E_\mathrm{F} - E_\mathrm{V}}{kT}\right), \tag{7.7}$$

with

$$N_\mathrm{C} = 2\left(\frac{m_\mathrm{e}\, kT}{2\pi\,\hbar^2}\right)^{3/2} \tag{7.8}$$

$$N_\mathrm{V} = 2\left(\frac{m_\mathrm{h}\, kT}{2\pi\,\hbar^2}\right)^{3/2}, \tag{7.9}$$

where N_C (N_V) is called the conduction-band (valence-band) edge density of states. The masses in (7.8) and (7.9) are the density of states masses given in (6.68) and (6.69). Values of $N_{\mathrm{C,V}}$ for Si, Ge and GaAs are given in Table 7.1.

Now, we assume that the Boltzmann approximation (Appendix E) can be used, i.e. the probability that a band state is populated is $\ll 1$. Then, the integral (7.1) can

Table 7.1 Band gap, intrinsic carrier concentration, conduction band and valence-band edge density of states at $T = 300\,\mathrm{K}$ for various semiconductors

	E_g (eV)	n_i (cm^{-3})	N_C (cm^{-3})	N_V (cm^{-3})
InSb	0.18	1.6×10^{16}		
InAs	0.36	8.6×10^{14}		
Ge	0.67	2.4×10^{13}	1.04×10^{19}	6.0×10^{18}
Si	1.124	1.0×10^{10}	7.28×10^{19}	1.05×10^{19}
GaAs	1.43	1.8×10^{6}	4.35×10^{17}	5.33×10^{18}
GaP	2.26	2.7×10^{0}		
GaN	3.3	$\ll 1$		

be executed analytically and the concentration n of electrons in the conduction band is given as

$$n = 2\left(\frac{m_e kT}{2\pi\hbar^2}\right)^{3/2} \exp\left(\frac{E_F - E_C}{kT}\right) = N_C \exp\left(\frac{E_F - E_C}{kT}\right). \tag{7.10}$$

For the Boltzmann approximation and a parabolic valence band, the density of holes is given by

$$p = 2\left(\frac{m_h kT}{2\pi\hbar^2}\right)^{3/2} \exp\left(-\frac{E_F - E_V}{kT}\right) = N_V \exp\left(-\frac{E_F - E_V}{kT}\right). \tag{7.11}$$

The product of the electron and hole density is

$$\begin{aligned} n\,p &= N_V\,N_C \exp\left(-\frac{E_C - E_V}{kT}\right) = N_V\,N_C \exp\left(-\frac{E_g}{kT}\right) \\ &= 4\left(\frac{kT}{2\pi\hbar^2}\right)^3 (m_{d,e}\,m_{d,h})^{3/2} \exp\left(-\frac{E_g}{kT}\right). \end{aligned} \tag{7.12}$$

Thus, the product np is *independent* of the position of the Fermi level, as long as the Boltzmann approximation is fulfilled, i.e. the Fermi level is not in the vicinity of one of the band edges within several kT:

$$E_V + 4kT < E_F < E_C - 4kT. \tag{7.13}$$

The relation (7.12) is called the mass-action law.

In Fig. 7.2, the product np is shown for silicon over a wide range of Fermi energies. If E_F is within the band gap, np is essentially constant. If the Fermi level is in the valence or conduction band, np decreases exponentially.

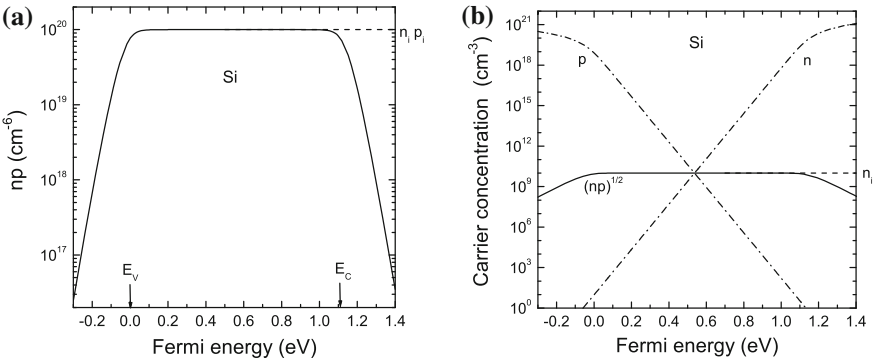

Fig. 7.2 (**a**) np for silicon at $T = 300\,\mathrm{K}$ as a function of the position of the Fermi level. The valence-band edge E_V is chosen as $E = 0$. np is constant for the range of Fermi energies given by (7.13) ($4kT \approx 0.1\,\mathrm{eV}$). (**b**) n, p and $\sqrt{np}$ as a function of the Fermi level

7.3 Intrinsic Conduction

First, we consider the conductivity of the intrinsic, i.e. an ideally pure, semiconductor. At $T = 0$ all electrons are in the valence band, the conduction band is empty and thus the conductivity is zero (a completely filled band cannot conduct current). Only at finite temperatures the electrons have a finite probability to be in a conduction-band state and to contribute to the conductivity. Due to neutrality, the electron and hole concentrations in the intrinsic semiconductors are the same, i.e. each electron in the conduction band comes from the valence band,

$$-n + p = 0, \tag{7.14}$$

or $n_\mathrm{i} = p_\mathrm{i}$. Therefore

$$\begin{aligned} n_\mathrm{i} = p_\mathrm{i} &= \sqrt{N_\mathrm{V} N_\mathrm{C}}\, \exp\left(-\frac{E_\mathrm{g}}{2kT}\right) \\ &= 2\left(\frac{kT}{2\pi\hbar^2}\right)^{3/2} (m_\mathrm{e} m_\mathrm{h})^{3/4}\, \exp\left(-\frac{E_\mathrm{g}}{2kT}\right). \end{aligned} \tag{7.15}$$

The mass-action law

$$np = n_\mathrm{i}\, p_\mathrm{i} = n_\mathrm{i}^2 = p_\mathrm{i}^2 \tag{7.16}$$

will be essential also for light and moderately doped semiconductors. The intrinsic carrier concentration is exponentially dependent on the band gap. Thus, in thermodynamic equilibrium intrinsic wide-gap semiconductors have much smaller electron concentrations than intrinsic small-gap semiconductors (see Table 7.1). The intrinsic carrier concentration of Si (in cm^{-3}) has been determined to be (within 1 %, T in K)

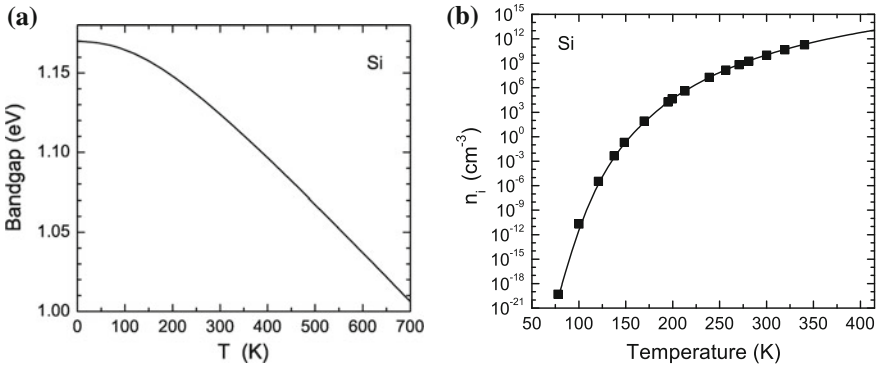

Fig. 7.3 (**a**) Band gap of silicon versus temperature. (**b**) Intrinsic carrier concentration of silicon versus temperature. *Solid line* is (7.17) using $E_g = 1.204\,\text{eV} - (2.73 \times 10^{-4}\,\text{eV/K})\,T$ [491], *symbols* are experimental data from [492]

$$n_i^{Si} = 1.640 \times 10^{15}\ T^{1.706}\ \exp\left(-\frac{E_g(T)}{2kT}\right) \tag{7.17}$$

for temperatures between 77 and 400 K [489, 490] (Fig. 7.3).

As we will see later in Part II, many semiconductor devices rely on regions of low conductivity (depletion layers) in which the carrier concentration is small. Since the carrier concentration cannot be smaller than the intrinsic concentration ($n + p \geq 2n_i$), an increase of temperature leads to increasing ohmic conduction in the depletion layers and thus to a reduction or failure of device performance. The small band gap of Ge leads to degradation of bipolar device performance already shortly above room temperature. For silicon, intrinsic conduction limits operation typically to temperatures below about 300 °C. For higher temperatures, as required for devices in harsh environments, such as close to motors or turbines, other semiconductors with wider band gaps need to be used, such as GaN, SiC or even diamond.

From the neutrality condition for the intrinsic semiconductor (7.14), (7.10) and (7.11), the Fermi level of the intrinsic semiconductor can be determined as

$$E_F = E_i = \frac{E_V + E_C}{2} + \frac{kT}{2}\ln\left(\frac{N_V}{N_C}\right) = \frac{E_V + E_C}{2} + \frac{3}{4}kT\ln\left(\frac{m_h}{m_e}\right). \tag{7.18}$$

Since the hole mass is perhaps a factor of ten larger than the electron mass, the second term has the order of kT. Thus, for typical semiconductors where $E_g \gg kT$, the intrinsic Fermi level, denoted by E_i, is close to the middle of the band gap, i.e. $E_i \approx (E_C + E_V)/2$.

The situation for an intrinsic semiconductor is schematically shown in Fig. 7.4b.

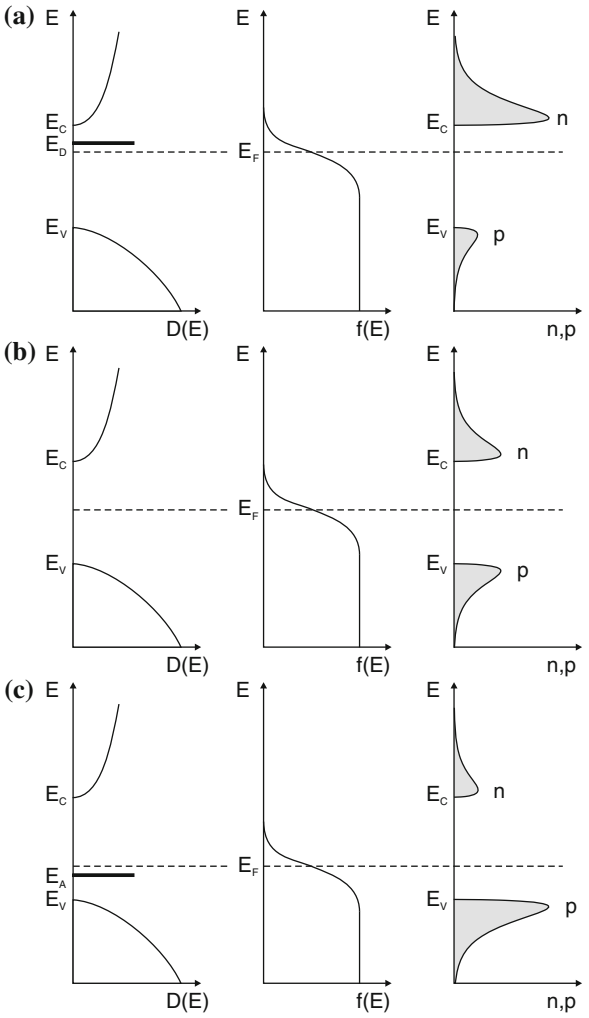

Fig. 7.4 Density of states (*left column*), Fermi distribution (*center column*) and carrier concentration (*right column*) for (**a**) n-type, (**b**) intrinsic and (**c**) p-type semiconductors in thermal equilibrium

7.4 Doping

7.4.1 Concept

The modification of the conductivity of a semiconductor using point defects is termed *doping*. In 1930 electrical conduction of semiconductors was attributed solely to impurities [493]. However 'chemically pure' substances become conductive upon deviation from stoichiometry, e.g. historically found for changes in the anion concentration and conductivity in CuI [31] (p-type) and ZnO [70] (n-type). The modification of CuI by exposure to different partial pressure of iodine in organic solutions

with different iodine concentration [34] and subsequently various concentrations of copper vacancies [494] can be considered the first doping of a semiconductor (1909).

The electronic levels of a defect or an impurity can exist within the forbidden gap of the bulk host material. These levels can lie close to the band edges or in the vicinity of the middle of the band gap. In a simplified approach, the first stem from *shallow* defects (Sect. 7.5), the latter from *deep* defects (Sect. 7.7).

7.4.2 Doping Principles

In [495] various doping principles are formulated. Essentially, the amount of impurities that lead to electrically active dopants is limited by the increasingly probable formation of compensating defects. In the case of donors, these are electron killers, e.g. n-type doping of Si:As is limited by the formation of V_{Si} [496]. In the case of acceptors, the compensating defects are hole killers. The so-called n-type *pinning energy* $E_F^{n,pin}$ is the Fermi level at which such killer defect (e.g. a cation vacancy) forms. When the Fermi level reaches the pinning energy, no further progress in n-type doping can be made, since the spontaneously generated electron killers will negate the introduced (impurity) donors. As a tendency, materials with low lying conduction band, i.e. large electron affinity (difference between vacuum level and conduction band) can be doped n-type. Similarly, p-type doping by acceptors, shifting the Fermi level towards the valence band, will meet at some point $E_F^{p,pin}$, called p-type pinning energy, when native hole killers, such as anion vacancies or cation interstitials form spontaneously. At this point, further p-type doping is no longer possible. p-type doping is facilitated by materials whose valance band maximum is close to the vacuum level [495].

A comparison of the wide gap materials ZnO, NiO and MgO is depicted in Fig. 7.5. The position of the pinning levels is marked on a common energy scale. From the position of $E_F^{n,pin}$ it can be understood, that ZnO can be highly n-doped while NiO

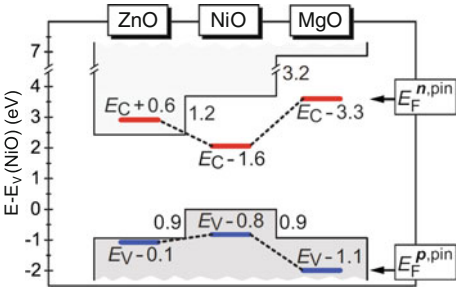

Fig. 7.5 Comparison of ZnO, NiO and MgO on a common energy scale, comparing conduction band and valence band edges and n-type (*red*) and p-type (*blue*) pinning energies (determined for metal-rich and oxygen-rich conditions, respectively). Adapted from [497]

and MgO cannot [497]. From $E_{\mathrm{F}}^{\mathrm{p,pin}}$, NiO can be doped p-type, while MgO cannot be doped at all.

For dopability, generally, it is also important that the ionized charges from impurities are free and thus contribute to the free charge carrier density and do not form localized states, e.g. due to polaronic effects (Sect. 8.6).

7.5 Shallow Defects

In Fig. 7.6, the positions of the energy levels of a variety of impurities are shown for Ge, Si and GaAs. An impurity for which the long-range Coulomb part of the ion-core potential determines the energetic level is termed a *shallow* impurity. The extension of the wavefunction is given by the Bohr radius. This situation is in contrast to a *deep* level where the short-range part of the potential determines the energy level. The extension of the wavefunction is then of the order of the lattice constant. A view on the history of the science of shallow impurity states is given in [498, 499].

We will consider first a group-IV semiconductor, Si, and (impurities) dopants from the groups III and V of the periodic system. When these are incorporated on a lattice site (with tetrahedral bonds), there is one electron too few (group III, e.g. B) or one electron too many (group V, e.g. As). The first case is called an *acceptor*, the latter a *donor.* The doping of III–V semiconductors is detailed in [501].

7.5.1 Donors

Silicon doped with arsenic is denoted as Si:As. The situation is schematically shown in Fig. 7.7. The arsenic atom has, after satisfying the tetrahedral bonds, an extra electron. This electron is bound to the arsenic atom via the Coulomb interaction since the ion core is positively charged compared to the silicon cores. If the electron is ionized, a fixed positive charge remains at the As site.

Without being in the silicon matrix, an arsenic atom has an ionization energy of 9.81 eV. However, in the solid the Coulomb interaction is screened by the dielectric constant of the material, typically ϵ_{r} is of the order of 10 for typical semiconductors. Additionally, the mass is renormalized (effective mass) by the periodic potential to a value that is smaller than the free electron mass. Within effective-mass theory (Appendix H) Bohr's theory of the hydrogen problem is scaled with the (isotropic) effective mass m_{e}^* and the dielectric constant ϵ_{r}, the binding energy (ionization energy) E_{D}^b of the electron to the shallow donor is (relative to the continuum given by the conduction-band edge E_{C})

$$E_{\mathrm{D}}^b = \frac{m_{\mathrm{e}}^*}{m_0}\,\frac{1}{\epsilon_{\mathrm{r}}^2}\,\frac{m_0\, e^4}{2\,(4\pi\epsilon_0\,\hbar)^2}. \tag{7.19}$$

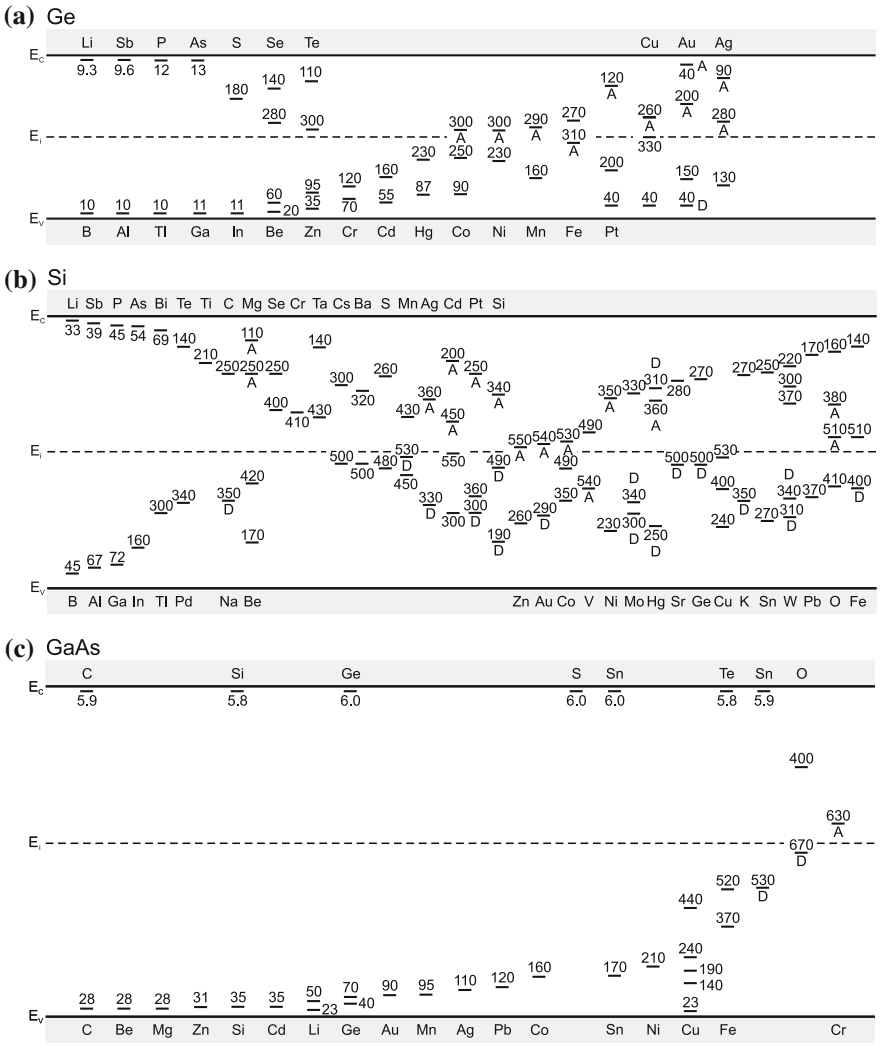

Fig. 7.6 Energetic position (ionization energy labeled in meV) of various impurities (A: acceptor, D: donor) in (**a**) Ge, (**b**) Si and (**c**) GaAs. Based on [500]

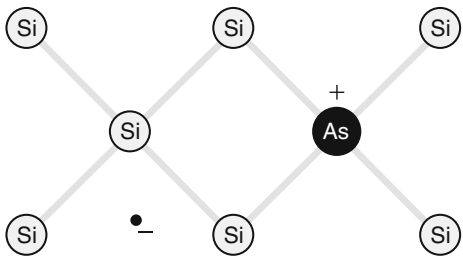

Fig. 7.7 Arsenic impurity in silicon. Arsenic donates one electron, and a fixed positive charge remains

The absolute energy position of the level is $E_D = E_C - E_D^b$. The first factor in the right side of (7.19) is the ratio of effective and free-electron mass, typically 1/10, the second factor is typically 1/100. The third factor is the ionization energy of the hydrogen atom, i.e. the Rydberg energy of 13.6 eV. Thus, the binding energy in the solid is drastically reduced by a factor of about 10^{-3} to the 10 meV regime. The excited states of the hydrogen-like spectrum can also be investigated experimentally (Sect. 9.7).

The extension of the wavefunction of the electron bound to the fixed ion is given by the Bohr radius

$$a_D = \frac{m_0}{m_e^*}\,\epsilon_r\,a_B, \tag{7.20}$$

where $a_B = 0.053$ nm denotes the hydrogen Bohr radius. For GaAs $a_D = 10.3$ nm. A similar value has been determined for InP [502]. For semiconductors with a non-isotropic band minimum, such as Si, Ge or GaP, an 'elliptically deformed' hydrogen problem with the masses m_l and m_t has to be treated [503].

An impurity that fulfills (7.19) is called an *effective-mass* impurity. For GaAs, the effective-mass donor has a binding energy of 5.715 meV, which is closely fulfilled for several chemical species (Table 7.3). In GaP, experimental values deviate considerably from the effective-mass donor (59 meV). For silicon, considering the anisotropic tensor of the effective masses, the result for the effective-mass donor binding energy is 29 meV [503]. Some experimentally observed values are summarized in Table 7.2.

Table 7.2 Binding energies E_D^b of Li and group-V donors in elemental semiconductors

	Li	N	P	As	Sb
C		1700	≈500		
Si	33		45	49	39
Ge	9.3		12.0	12.7	9.6

Data for carbon from [505]. All values in meV

Table 7.3 Binding energies E_D^b of donors in GaAs (data from [506]), GaP (data from [507]) and GaN (low concentration limits, data from [508, 509])

	V site		III site	
GaAs	S	5.854	C	5.913
	Se	5.816	Si	5.801
	Te	5.786	Ge	5.937
GaP	O	897	Si	85
	S	107	Ge	204
	Se	105	Sn	72
	Te	93		
GaN	O	39	Si	22
			Ge	19

All values in meV

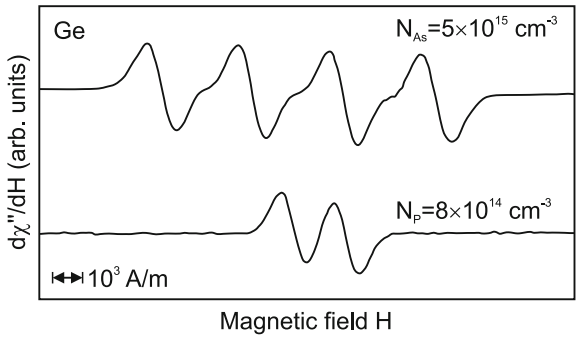

Fig. 7.8 Electron spin resonance signal from As and P in Ge with the magnetic field **H** parallel to [100], $T \approx 1.3\,\mathrm{K}$. Adapted from [504]

Deviations from the effective-mass theory are due to modification of the potential in the immediate vicinity of the impurity atom and breakdown of the effective-mass formalism.

Different impurities can have quite similar binding energies. They can be distinguished, e.g. by electron spin resonance (ESR). At low temperatures the electron is localized on the impurity and the hyperfine interaction with the nucleus can be resolved in ESR. In Fig. 7.8 data are shown for As and P in germanium. The multiplets distinguish the nuclear spins $I = 3/2$ for arsenic (^{75}As) and $I = 1/2$ for phosphorus (^{31}P) [504].

The donors are typically distributed statistically (randomly) in the solid. Otherwise their distribution is called clustered. The concentration of donors is labeled N_D and usually given in cm^{-3}.

The concentration of donors populated with an electron (neutral donors) is denoted by N_D^0, the concentration of ionized donors (positively charged) is N_D^+. Other conventions in the literature label the concentrations N_1 and N_0, respectively:

$$N_1 = N_\mathrm{D}^0 = N_\mathrm{D}\, f_\mathrm{e}(E_\mathrm{D}) \tag{7.21a}$$

$$N_0 = N_\mathrm{D}^+ = N_\mathrm{D}\,(1 - f_\mathrm{e}(E_\mathrm{D})), \tag{7.21b}$$

with $f_\mathrm{e}(E_\mathrm{D}) = \left[1 + \exp(E_\mathrm{D} - E_\mathrm{F})\right]^{-1}$. For the sum of these quantities the condition

$$N_\mathrm{D} = N_\mathrm{D}^+ + N_\mathrm{D}^0 \tag{7.22}$$

holds.

The ratio of the two concentrations is first given as (caveat: this formula will be modified below)

$$\frac{N_\mathrm{D}^0}{N_\mathrm{D}^+} = \frac{N_1}{N_0} = \frac{f}{1-f} = \exp\left(\frac{E_\mathrm{F} - E_\mathrm{D}}{kT}\right). \tag{7.23}$$

Now, the degeneracy of the states has to be considered. The donor charged with one electron has a 2-fold degeneracy $g_1 = 2$ since the electron can take the spin up and

down states. The degeneracy of the ionized (empty) donor is $g_0 = 1$. Additionally, we assume here that the donor cannot be charged with a second electron (cmp. Sect. 7.7.2). Due to Coulomb interaction, the energy level of the possible N_D^- state is in the conduction band. Otherwise, a multiply charged center would be present. We also do not consider excited states of N_D^0 that might be in the band gap as well. In the following, we will continue with $\hat{g}_D = g_1/g_0 = 2$ as suggested in [510].[2] We note that the definition of the degeneracy factor for donors (and acceptors, see (7.39)) is not consistent in the literature as summarized in [511]. Considering now the degeneracy, (7.23) is modified to

$$\frac{N_D^0}{N_D^+} = \frac{N_1}{N_0} = \hat{g}_D \exp\left(\frac{E_F - E_D}{kT}\right). \tag{7.24}$$

This can be understood from thermodynamics (cf. Sect. 4.2.2), a rate analysis or simply the limit $T \to \infty$.

The probabilities f^1 and f^0 for a populated or empty donor, respectively, are

$$f^1 = \frac{N_1}{N_D} = \frac{1}{\hat{g}_D^{-1} \exp\left(\frac{E_D - E_F}{kT}\right) + 1} \tag{7.25a}$$

$$f^0 = \frac{N_0}{N_D} = \frac{1}{\hat{g}_D \exp\left(-\frac{E_D - E_F}{kT}\right) + 1}. \tag{7.25b}$$

First, we assume that no carriers in the conduction band stem from the valence band (no intrinsic conduction). This will be the case at sufficiently low temperatures when $N_D \gg n_i$. Then the number of electrons in the conduction band is equal to the number of ionized donors, i.e.

$$n = f^0 N_D = N_0 = \frac{N_D}{1 + \hat{g}_D \exp\left(\frac{E_F - E_D}{kT}\right)} = \frac{1}{1 + n/n_1} N_D,$$

with $n_1 = (N_C/\hat{g}_D) \exp\left(-E_D^b/kT\right)$. The neutrality condition (its general from is given in (7.41)) is

$$-n + N_D^+ = -n + N_0 = 0, \tag{7.26}$$

leading to the equation (n is given by (7.10))

$$N_C \exp\left(\frac{E_F - E_C}{kT}\right) - \frac{N_D}{1 + \hat{g} \exp\left(\frac{E_F - E_D}{kT}\right)} = 0. \tag{7.27}$$

[2] We do not agree with the treatment of the conduction band valley degeneracy in [510] for the donor degeneracy factor for Ge and Si.

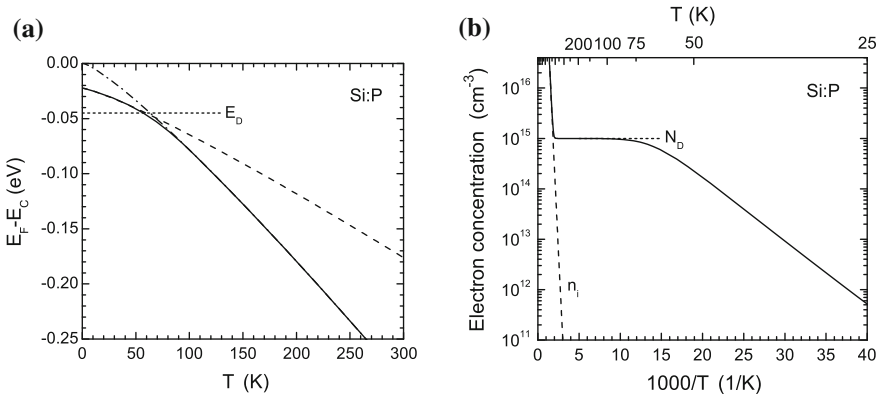

Fig. 7.9 (**a**) Position of the Fermi level in Si:P ($N_D = 10^{15}\,\mathrm{cm}^{-3}$, $E_D^b = 45\,\mathrm{meV}$, no acceptors) as a function of temperature *without* consideration of intrinsic carriers. Zero energy refers to the (temperature-dependent, Table 6.4) conduction-band edge E_C with approximative solutions for low (*dashed line*, (7.29)) and high (*dash-dotted line*, (7.30)) temperatures. (**b**) Corresponding density of conduction-band electrons as a function of temperature

Solving this equation will yield the Fermi level (as a function of temperature T, doping level E_D and doping concentration N_D).[3] The solution is

$$E_F = E_C - E_D^b + kT \ln\left(\frac{\left[1 + 4\hat{g}_D \frac{N_D}{N_C} \exp\left(\frac{E_D^b}{kT}\right)\right]^{1/2} - 1}{2\,\hat{g}_D}\right). \tag{7.28}$$

For $T \to 0$ the Fermi level is, as expected, in the center between the populated and unpopulated states, i.e. at $E_F = E_C - E_D^b/2$. In Fig. 7.9a the position of the Fermi is shown for a donor with 45 meV binding energy in Si. For low temperatures the solution can be approximated as (dashed curve in Fig. 7.9b)

$$E_F \cong E_C - \frac{1}{2} E_D^b + \frac{1}{2} kT \ln\left(\frac{N_D}{\hat{g}_D N_C}\right). \tag{7.29}$$

The freeze-out of carriers in n-type silicon has been discussed in detail in [512], taking into account the effects of the fine structure of the donor states. We note that the fairly high donor binding energy in silicon leads to freeze-out of carriers at about 40 K and is thus limiting for the low-temperature performance of devices. Ge has smaller donor ionization energies and subsequently a lower freeze-out temperature of 20 K. For n-type GaAs, conductivity is preserved down to even lower temperatures.

[3] As usual, the Fermi level is determined by the global charge neutrality, see also Sect. 4.2.2.

We note that the freeze-out of carriers involves the recombination of free electrons with the ionized donors. This aspect is considered in Sect. 10.9. Microscopically this process is equal to the emission of a (far infrared) photon [513, 514]. Similarly the release of an electron from the donor is due to the absorption of a photon.

For higher temperatures, when the electron density saturates towards N_D, the approximate solution is (dash-dotted curve in Fig. 7.9a)

$$E_F \cong E_C + kT \ln\left(\frac{N_D}{N_C}\right). \tag{7.30}$$

The electron density n is given (still in the Boltzmann approximation) by

$$\begin{aligned} n &= N_C \exp\left(-\frac{E_D^b}{kT}\right) \frac{\left[1 + 4\,\hat{g}_D \frac{N_D}{N_C} \exp\left(\frac{E_D^b}{kT}\right)\right]^{1/2} - 1}{2\,\hat{g}_D} \\ &= \frac{2\,N_D}{1 + \left[1 + 4\,\hat{g}_D \frac{N_D}{N_C} \exp\left(\frac{E_D^b}{kT}\right)\right]^{1/2}}. \end{aligned} \tag{7.31}$$

The theoretical electron density as a function of temperature is shown in Fig. 7.9b. It fits very well to experimental data for arsenic doped germanium [515] as shown in Fig. 7.10 (Arrhenius plot, $\ln n$ vs. $1/T$).

For low temperatures, the solution (7.31) is close to

$$n \cong \sqrt{\frac{N_D N_C}{\hat{g}_D}} \exp\left(-\frac{E_D^b}{2kT}\right) = \sqrt{n_1\, N_D}. \tag{7.32}$$

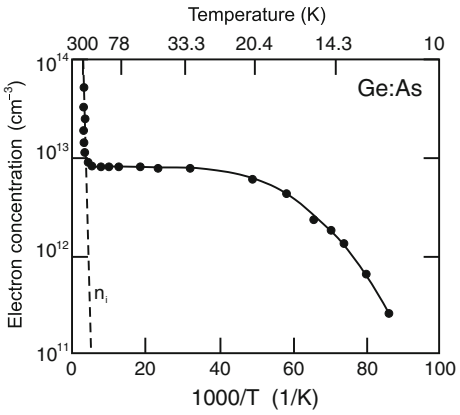

Fig. 7.10 Electron concentration as a function of temperature for a Ge:As sample with $N_D \approx 10^{13}\,\mathrm{cm}^{-3}$. *Solid line* is fit to the data with a donor binding energy of 12.7 meV. Adapted from [515]

For high temperatures, $n \cong N_\mathrm{D}$. This regime is called *exhaustion* or *saturation* since all possible electrons have been ionized from their donors. We note that even in this case $np = n_\mathrm{i} p_\mathrm{i}$ holds, however, $n \gg p$.

While the characteristic energy for the ionization of electrons from donors is E_D^b, at high enough temperatures electrons are transferred also from the valence band into the conduction band. Thus, in order to make the above consideration valid for all temperatures, the intrinsic conduction also has to be considered. The neutrality condition (still in the absence of any acceptors) is

$$-n + p + N_\mathrm{D}^+ = 0. \tag{7.33}$$

Using (7.10) and $p = n_\mathrm{i}^2/n$, the equation reads:

$$N_\mathrm{C} \exp\left(\frac{E_\mathrm{F} - E_\mathrm{C}}{kT}\right) - \frac{n_\mathrm{i}^2}{N_\mathrm{C} \exp\left(\frac{E_\mathrm{F}-E_\mathrm{C}}{kT}\right)} - \frac{N_\mathrm{D}}{1 + \hat{g}_\mathrm{D} \exp\left(\frac{E_\mathrm{F}-E_\mathrm{D}}{kT}\right)} = 0. \tag{7.34}$$

The solution is

$$E_\mathrm{F} = E_\mathrm{C} - E_\mathrm{D}^b + kT \ln\left[\frac{\beta/\gamma + \gamma/N_\mathrm{C}^2 - 1}{3\,\hat{g}_\mathrm{D}}\right], \tag{7.35}$$

with

$$\gamma = \left(\frac{-N_\mathrm{C}^4\,\alpha + \sqrt{(N_\mathrm{C}^4\,\alpha)^2 - 4(N_\mathrm{C}^2\,\beta)^3}}{2}\right)^{1/3} \tag{7.36a}$$

$$\beta = N_\mathrm{C}^2 + 3\hat{g}_\mathrm{D} N_\mathrm{C} N_\mathrm{D} \exp\left(\frac{E_\mathrm{D}}{kT}\right) + 3\hat{g}_\mathrm{D}^2 n_\mathrm{i}^2 \exp\left(\frac{2E_\mathrm{D}}{kT}\right) \tag{7.36b}$$

$$\alpha = 2N_\mathrm{C}^2 + 9\hat{g}_\mathrm{D} N_\mathrm{C} N_\mathrm{D} \exp\left(\frac{E_\mathrm{D}}{kT}\right) - 18\hat{g}_\mathrm{D}^2 n_\mathrm{i}^2 \exp\left(\frac{2E_\mathrm{D}}{kT}\right). \tag{7.36c}$$

The temperature-dependent position of the Fermi level is shown in Fig. 7.11. The carrier concentration is given by

$$n = N_\mathrm{C} \exp\left(-\frac{E_\mathrm{D}^b}{kT}\right) \frac{\beta/\gamma + \gamma/N_\mathrm{C}^2 - 1}{3\,\hat{g}_\mathrm{D}}. \tag{7.37}$$

The three important regimes are the intrinsic conduction at high temperatures when $n_\mathrm{i} \gg N_\mathrm{D}$, the exhaustion at intermediate temperatures when $n_\mathrm{i} \ll N_\mathrm{D}$ and $kT > E_\mathrm{D}$, and finally the freeze-out regime for $kT \ll E_\mathrm{D}$ at low temperatures when the electrons condense back into the donors. The three regimes can be seen in the experimental data of Fig. 7.10 for donors (n-Ge) and Fig. 7.15 for acceptors (p-Ge).

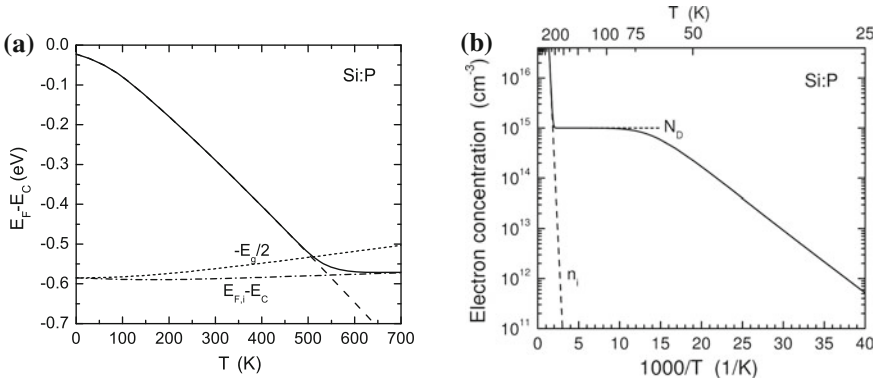

Fig. 7.11 (**a**) Position of the Fermi level in Si:P ($N_D = 10^{15}\,\mathrm{cm}^{-3}$, $E_D^b = 45\,\mathrm{meV}$, no acceptors) as a function of temperature. The temperature dependence of the band gap (as given in Table 6.4) has been taken into account. Zero energy refers to the conduction-band edge for all temperatures. The *dotted curve* shows $E_g/2$. The *dashed* (*dash-dotted*) *line* shows the low- (high-) temperature limit according to (7.29) and (7.18), respectively. (**b**) Corresponding electron concentration as a function of temperature. The *dashed line* shows the intrinsic carrier density

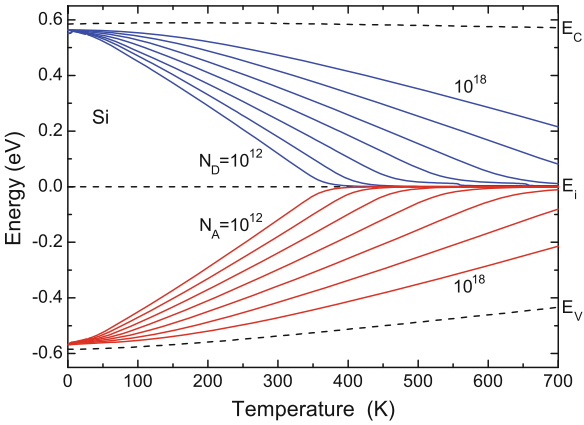

Fig. 7.12 Fermi level in silicon as a function of temperature for various doping levels (n-type (*blue lines*) and p-type (*red lines*)) of $10^{12}, 10^{13}, \ldots, 10^{18}\,\mathrm{cm}^{-3}$. The intrinsic Fermi level is chosen as zero energy for all temperatures. The (temperature-dependent) conduction and valence band edges are shown as *dashed lines*

A similar plot as in Fig. 7.11a is shown in Fig. 7.12. With increasing temperature the Fermi level shifts from close to the band edge towards the band center. At higher doping the shift begins at higher temperatures.

The electronic states of individual donors can be directly visualized by scanning tunneling microscopy (STM) as shown in Fig. 7.13 for Si:P. For small negative bias, tunneling occurs through the charged dopant that is located within the first three monolayers. At high negative bias the large contribution from the filled valence band masks the effect of the donor. This image, however, shows that the contrast attributed to the dopant atom is not due to surface defects or absorbates.

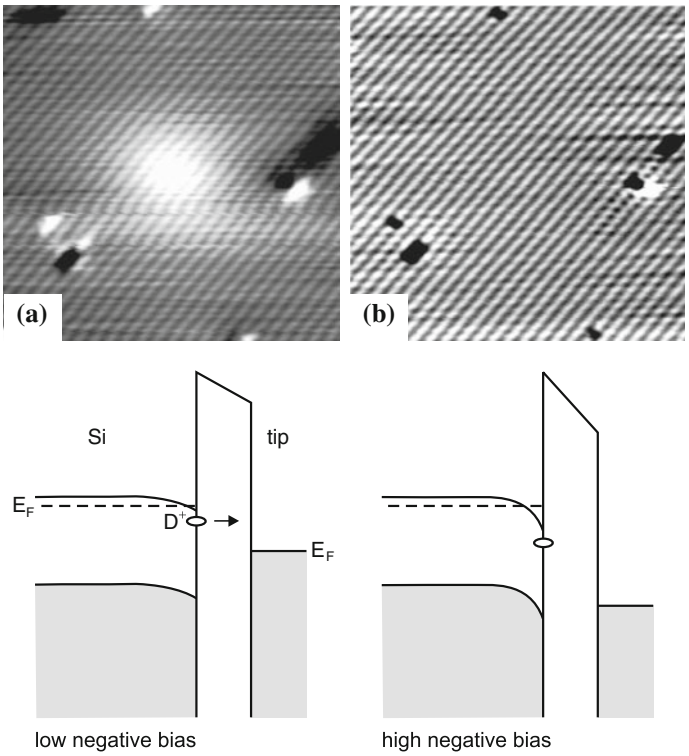

Fig. 7.13 Filled-state image of a phosphorus atom underneath a Si (001) surface at a tunneling current of 110 pA. The doping level is $5 \times 10^{17}\,\mathrm{cm}^{-3}$. (**a**) Sample bias -0.6 V, (**b**) sample bias -1.5 V between Si:P and tip. Image sizes are $22 \times 22\,\mathrm{nm}^2$. Reprinted with permission from [516], © 2004 APS. *Lower row* Schematic band diagrams for the two bias situations

7.5.2 Acceptors

A group-III atom in Si has one electron too few for the tetrahedral bond. Thus, it 'borrows' an electron from the electron gas (in the valence band) and thus leaves a missing electron (termed hole) in the valence band (Fig. 7.14). The energy level of the impurity is in the gap close to the valence-band edge. The latter consideration is made in the electron picture. In the hole picture, the acceptor ion has a hole and the hole ionizes (at sufficient temperature) into the valence band. After ionization the acceptor is charged negatively. Also, for this system a Bohr-like situation arises that is, however, more complicated than for donors because of the degeneracy of the valence bands and their warping.

In Table 7.4 the acceptor binding energies E_A^b for group-III atoms in C, Ge and Si are listed. The absolute acceptor energy is given as $E_\mathrm{A} = E_\mathrm{V} + E_\mathrm{A}^b$. In Table 7.5 acceptor binding energies are listed for GaAs, GaP and GaN. While in GaAs some

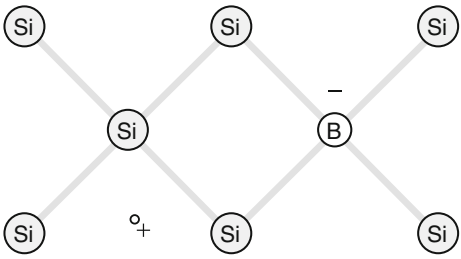

Fig. 7.14 Boron impurity in silicon. Boron accepts one electron and a fixed negative charge remains

Table 7.4 Binding energies E_A^b of group-III acceptors in elemental semiconductors

	B	Al	Ga	In
C	369			
Si	45	57	65	16
Ge	10.4	10.2	10.8	11.2

Data for diamond from [517, 518]. All values in meV

Table 7.5 Binding energies E_A^b of acceptors in GaAs, GaP and GaN (low concentration values, data from [519, 520])

	V site		III site	
GaAs	C	27	Be	28
	Si	34.8	Mg	28.8
	Ge	40.4	Zn	30.7
	Sn	167	Cd	34.7
GaP	C	54	Be	57
	Si	210	Mg	60
	Ge	265	Zn	70
			Cd	102
GaN	C	230	Mg	220
	Si	224	Zn	340
			Cd	550

All values in meV

acceptors are close to the effective mass value of 27 meV, in GaP the deviation from the effective-mass value ≈50 meV is large.

When the conductivity is determined by holes or electrons, the material is called p-type or n-type, respectively. We note that some metals also show hole conduction (e.g. Al). However, for metals the conductivity type is fixed, while the same semiconductor can be made n- or p-type with the appropriate doping.

The acceptor concentration is denoted by N_A. The concentration of neutral acceptors is N_A^0, the concentration of charged acceptors is N_A^-. Of course

$$N_A = N_A^0 + N_A^-. \tag{7.38}$$

The ratio of the degeneracy of the (singly) filled and empty acceptor level is $\hat{g}_A$. In Ge $\hat{g}_A = 4$ since the localized hole wave function may be formed in EMA with four Bloch wave functions (heavy and light holes) [521]. For Si with its small split-off energy (Table 6.6) $\hat{g}_A = 6$ according to [522]. For doubly ionized acceptors, e.g. Zn in Si and Ge (see Sect. 7.7.3), the more shallow level ($Zn^- \rightarrow Zn^0$) has $\hat{g}_A = 6/4 = 1.5$ in Ge [522]. A more general discussion of the degeneracy factor for multiply charged acceptors can be found in [510, 523]. Similar to the considerations for electrons and donors we have

$$\frac{N_A^0}{N_A^-} = \hat{g}_A \exp\left(-\frac{E_F - E_A}{kT}\right). \tag{7.39}$$

The population of the acceptor levels is given by

$$N_A^- = \frac{N_A}{1 + \hat{g}_A \exp\left(-\frac{E_F - E_A}{kT}\right)}. \tag{7.40}$$

The formulas for the position of the Fermi level and the hole density are analogous to those obtained for electrons and donors and will not be explicitly given here. The analogue to Fig. 7.11b is shown for data on p-doped Ge [524, 525] in Fig. 7.15. The acceptor activation energy is 11 meV which could be due to various impurities (cf. Table 7.4). The different impurities (B, Al, Ga) can be distinguished by photothermal ionization spectroscopy [525] (cmp. Sect. 9.7).

In Fig. 7.12, the temperature dependence of the Fermi level is included for p-type Si. With increasing temperature the Fermi level shifts from the valence-band edge (For $T = 0$, $E_F = E_V + E_A^b/2$) towards the middle of the band gap (intrinsic Fermi level).

Also, the wavefunction at acceptors can be imaged using scanning tunneling microscopy [526]. In [527] images of ionized and neutral Mn in GaAs have been

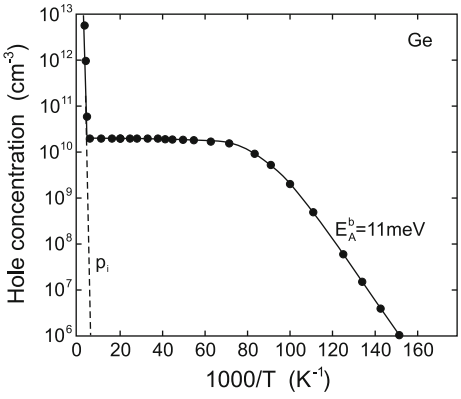

Fig. 7.15 Carrier concentration as a function of temperature for p-type Ge. The net shallow level concentration is $2 \times 10^{10}\,cm^{-3}$. *Solid line* is fit to the data, the *dashed line* indicates the intrinsic hole concentration p_i. Adapted from [525]

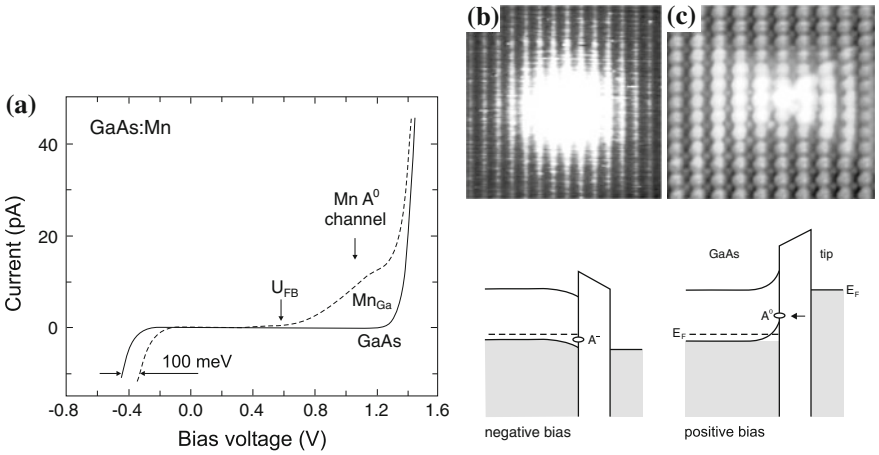

Fig. 7.16 (**a**) Tunneling I–V characteristic of GaAs:Mn sample. *Solid* (*dashed*) line is for pure GaAs (subsurface Mn on Ga site). U_{FB} denotes the simulated flat-band voltage. Adapted from [527]. (**b**, **c**) STM images of a Mn atom underneath a GaAs (110) surface. The doping level is $3 \times 10^{18}\,\mathrm{cm}^{-3}$. (**b**) Sample bias $-0.7\,\mathrm{V}$, (**c**) sample bias $+0.6\,\mathrm{V}$. Below the images are schematic band diagrams of GaAs:Mn and tip. Image sizes are (**b**) $8 \times 8\,\mathrm{nm}^2$ and (**c**) $5.6 \times 5\,\mathrm{nm}^2$. Reprinted with permission from [527], ©2004 APS. *Lower row* under parts (**a**, **b**): Schematic band diagrams for the two bias situations

reported (Fig. 7.16b). The tunneling I–V characteristics are shown in Fig. 7.16a. At negative bias, the acceptor is ionized and appears spherically symmetric due to the effect of the A^- ion Coulomb potential on the valence-band states. At intermediate positive voltages, tunneling is through the neutral state. The wavefunction of A^0 looks like a bow-tie due to the admixture of d-wavefunctions [528]. The Mn atom is presumably in the third subsurface atomic layer. At even higher positive bias the contrast due to the dopant is lost because the image is dominated by a large tunneling current from the tip to the empty conduction band.

7.5.3 Compensation

When donors *and* acceptors are simultaneously present, some of the impurities will compensate each other. Electrons from donors will recombine with holes on the acceptors. Depending on the quantitative situation the semiconductor can be n- or p-type. This situation can be invoked by intentional doping with donors or acceptors or by the unintentional background of donors (acceptors) in p-doped (n-doped) material. Also the formation of pairs, exhibiting a new defect level different from the single donor or single acceptor, has been described, e.g. for Se and B in silicon [273].

The charge-neutrality condition (now finally in its most general form) reads

$$-n + p - N_{\mathrm{A}}^{-} + N_{\mathrm{D}}^{+} = 0. \tag{7.41}$$

We will now discuss the case of the presence of donors and acceptors, but limit ourselves to sufficiently low temperatures (or wide band gaps) such that the intrinsic carrier density can be neglected. We assume Boltzmann statistics and assume here $N_{\mathrm{D}} > N_{\mathrm{A}}$. Then it is a very good approximation to use $N_{\mathrm{A}}^{-} = N_{\mathrm{A}}$ since there are enough electrons from the donors to recombine with (and thus compensate) all acceptors. Under the given assumptions regarding the temperature $p = 0$ and the material is n-type. Thus, in order to determine the position of the Fermi level, the charge-neutrality condition

$$n + N_{\mathrm{A}} - N_{\mathrm{D}}^{+} = 0 \tag{7.42}$$

must be solved (compare to (7.27))

$$N_{\mathrm{C}} \exp\left(\frac{E_{\mathrm{F}} - E_{\mathrm{C}}}{kT}\right) + N_{\mathrm{A}} - \frac{N_{\mathrm{D}}}{1 + \hat{g} \exp(\frac{E_{\mathrm{F}} - E_{\mathrm{D}}}{kT})} = 0. \tag{7.43}$$

We rewrite (7.42) and find $N_{\mathrm{D}} - N_{\mathrm{A}} - n = N_{\mathrm{D}}^{0} = N_{\mathrm{D}}^{+}\, \hat{g}_{\mathrm{D}} \exp\left(\frac{E_{\mathrm{F}} - E_{\mathrm{D}}}{kT}\right)$ using (7.24). Using again (7.42) and also (7.10), (7.43) can be written as

$$\frac{n\,(n + N_{\mathrm{A}})}{N_{\mathrm{D}} - N_{\mathrm{A}} - n} = \frac{N_{\mathrm{C}}}{\hat{g}_{\mathrm{D}}} \exp\left(-\frac{E_{\mathrm{D}}^{\mathrm{b}}}{kT}\right), \tag{7.44}$$

a form given in [529]. Analogously for compensated p-type material

$$\frac{p\,(p + N_{\mathrm{D}})}{N_{\mathrm{A}} - N_{\mathrm{D}} - p} = \frac{N_{\mathrm{V}}}{\hat{g}_{\mathrm{A}}} \exp\left(-\frac{E_{\mathrm{A}}^{\mathrm{b}}}{kT}\right) \tag{7.45}$$

holds.

The solution of (7.43) is

$$E_{\mathrm{F}} = E_{\mathrm{C}} - E_{\mathrm{D}}^{b} + kT \ln\left(\frac{\left[\alpha^2 + 4\hat{g}_{\mathrm{D}} \frac{N_{\mathrm{D}} - N_{\mathrm{A}}}{N_{\mathrm{C}}} \exp\left(\frac{E_{\mathrm{D}}^{b}}{kT}\right)\right]^{1/2} - \alpha}{2\,\hat{g}_{\mathrm{D}}}\right), \tag{7.46}$$

with

$$\alpha = 1 + \hat{g}_{\mathrm{D}} \frac{N_{\mathrm{A}}}{N_{\mathrm{C}}} \exp\left(\frac{E_{\mathrm{D}}^{b}}{kT}\right) = 1 + \frac{N_{\mathrm{A}}}{\beta} \tag{7.47a}$$

$$\beta = \frac{N_\mathrm{C}}{\hat{g}_\mathrm{D}} \exp\left(-\frac{E_\mathrm{D}^\mathrm{b}}{kT}\right). \tag{7.47b}$$

The carrier density is best obtained from (7.44),

$$n = \sqrt{(N_\mathrm{A} - \beta)^2 + 4\,N_\mathrm{D}\,\beta} - \frac{N_\mathrm{A} + \beta}{2}. \tag{7.48}$$

For $N_\mathrm{A} = 0$ we have $\alpha = 1$ and (7.28) is reproduced, as expected. For $T = 0$ (and $N_\mathrm{A} \neq 0$) the Fermi energy lies at $E_\mathrm{F} = E_\mathrm{D}$ since the donor level is *partially* filled ($N_\mathrm{D}^0 = N_\mathrm{D} - N_\mathrm{A}$). For low temperatures the Fermi level is approximated by

$$E_\mathrm{F} \cong E_\mathrm{C} - E_\mathrm{D}^b + kT\,\ln\left(\frac{N_\mathrm{D}/N_\mathrm{A} - 1}{\hat{g}_\mathrm{D}}\right). \tag{7.49}$$

The corresponding carrier density at low temperatures is

$$n = \frac{N_\mathrm{C}}{\hat{g}_\mathrm{D}} \exp\left(-\frac{E_\mathrm{D}^b}{kT}\right)\left(\frac{N_\mathrm{D}}{N_\mathrm{A}} - 1\right). \tag{7.50}$$

For higher temperatures (7.32) holds approximately for $n > N_\mathrm{A}$; the slope is now given by $E_\mathrm{D}^b/2$ as in the uncompensated case (Fig. 7.17b). For sufficiently high

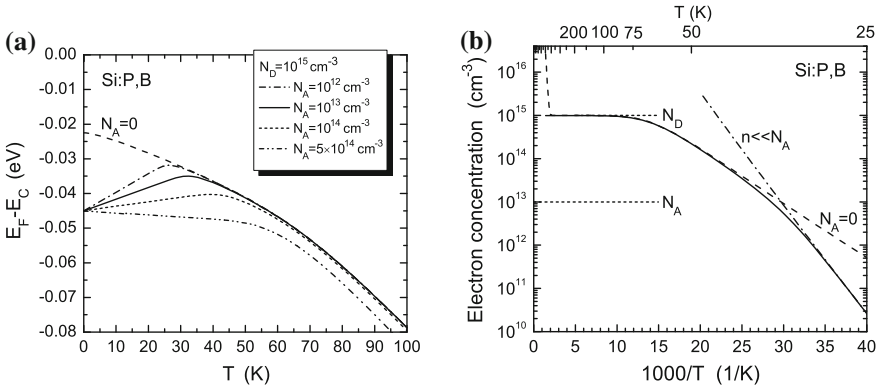

Fig. 7.17 (**a**) Position of Fermi level in partially compensated Si:P,B ($N_\mathrm{D} = 10^{15}\,\mathrm{cm}^{-3}$, $E_\mathrm{D}^b = 45\,\mathrm{meV}$, $E_\mathrm{A}^b = 45\,\mathrm{meV}$, *solid line* $N_\mathrm{A} = 10^{13}\,\mathrm{cm}^{-3}$, *dashed line* $N_\mathrm{A} = 0$, *dash-dotted line* $N_\mathrm{A} = 10^{12}\,\mathrm{cm}^{-3}$, *short-dashed line* $N_\mathrm{A} = 10^{14}\,\mathrm{cm}^{-3}$, *dash-double dotted line* $N_\mathrm{A} = 5 \times 10^{14}\,\mathrm{cm}^{-3}$) as a function of temperature. (**b**) Corresponding electron concentration for $N_\mathrm{A} = 10^{13}\,\mathrm{cm}^{-3}$ as a function of temperature (neglecting intrinsic carriers), *dashed line* for $N_\mathrm{A} = 0$ according to (7.31), *dash-dotted line* approximation for $n \ll N_\mathrm{A}$ as in (7.50)

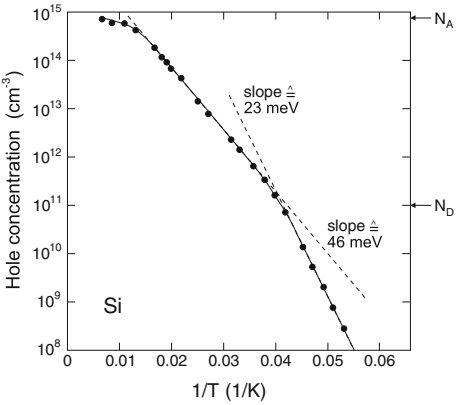

Fig. 7.18 Hole density in p-type silicon ($N_A = 7.4 \times 10^{14}\,cm^{-3}$, $E_A^b = 46\,meV$ (probably boron) and partial compensation with $N_D = 1.0 \times 10^{11}\,cm^{-3}$). Adapted from [530]

temperatures in the exhaustion regime (but still $n_i < n$) the electron density is given by

$$n \cong N_D - N_A. \tag{7.51}$$

At even higher temperatures the electron density will be determined by the intrinsic carrier concentration. Only in this case $p \neq 0$!

An experimental example is shown in Fig. 7.18 for partially compensated p-Si (with $N_D \ll N_A$). The change of slope around $p \approx N_D$ is obvious.

If donors are added to a p-type semiconductor, first the semiconductor remains p-conducting as long as $N_D \ll N_A$. If the donor concentration becomes larger than the acceptor concentration, the conductivity type switches from p- to n-conduction. If the impurities are exhausted at room temperature, the lowest carrier concentration is reached for $N_D = N_A$. Such a scenario is shown for p-type $In_xGa_{1-x}As_{1-y}N_y$ doped with various concentrations of Si in Fig. 7.19. At high Si incorporation, the number of charge carriers saturates due to autocompensation (see Sect. 7.5.5) and the formation of Si precipitates. Since the ionization energies of donors and acceptors are typically different, the situation for $N_D \approx N_A$ needs, in general, to be investigated carefully and will depend on the temperature.

7.5.4 Multiple Impurities

If more than one donor species is present, (7.43) can be generalized, e.g. for the case of two donors D1 and D2 in the presence of compensating acceptors,

$$n + N_A - \frac{N_{D1}}{1 + \hat{g}_1 \exp(\frac{E_F - E_{D1}}{kT})} - \frac{N_{D2}}{1 + \hat{g}_2 \exp(\frac{E_F - E_{D2}}{kT})} = 0. \tag{7.52}$$

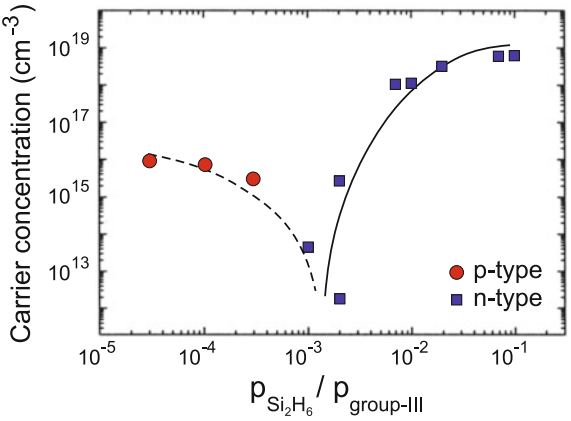

Fig. 7.19 Carrier concentration and conductivity type (*red circles* p, *blue squares* n) for MOVPE-grown $In_xGa_{1-x}As_{1-y}N_y$ layers on GaAs (001) (layer thickness $\approx 1\,\mu m$, $x \approx 5\%$, $y \approx 1.6\%$) doped with different amounts of silicon. The ordinate is the ratio of the partial pressures of disilane and the group-III precursors (TMIn and TMGa) in the gas phase entering the MOVPE reactor. *Lines* are guides to the eye. Experimental data from [531]

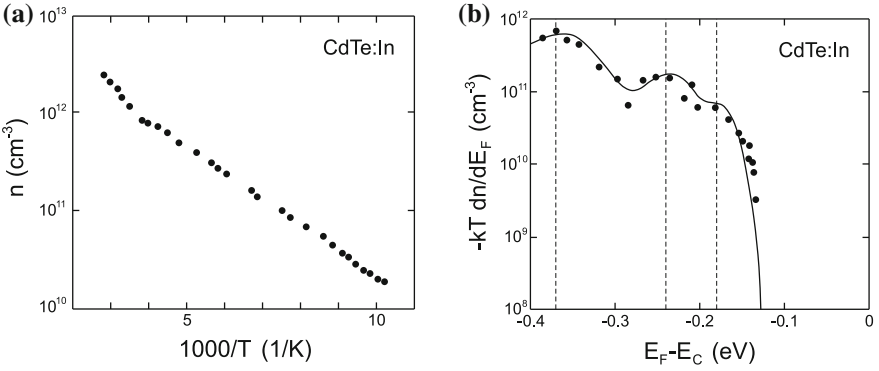

Fig. 7.20 (**a**) Electron concentration versus temperature as determined from Hall effect for a CdTe sample doped with indium. (**b**) $-kT\,dn/dE_F$, as determined from the experimental Hall data (*symbols*). The *solid line* is theory for three donor levels ($E_{D1} = E_C - 0.37\,eV$, $N_{D1} = 2.5 \times 10^{12}\,cm^{-3}$; $E_{D2} = E_C - 0.24\,eV$, $N_{D2} = 7.0 \times 10^{11}\,cm^{-3}$; $E_{D3} = E_C - 0.18\,eV$, $N_{D3} = 2.5 \times 10^{11}\,cm^{-3}$) whose energy positions are indicated by *dashed lines*. Adapted from [533]

This case is treated in [532]. Simple high and low temperature approximations can be found where the trap with the larger and smaller activation energy, respectively, dominates. The case for multiple acceptors (and compensating donors) is treated analogously. As detailed in [533], the function dn/dE_F has a maximum at the donor level position; this can be used to visualize the contribution of several donors (with sufficiently different binding energies) from $n(T)$ as measured by Hall effect (Fig. 7.20).

7.5.5 *Amphoteric Impurities*

If an impurity atom can act as a donor and acceptor it is called amphoteric. This can occur if the impurity has several levels in the band gap (such as Au in Ge or Si). In this case, the nature of the impurity depends on the position of the Fermi level. Another possibility is the incorporation on different lattice sites. For example, carbon in GaAs is a donor if incorporated on the Ga-site. On the As-site carbon acts as an acceptor.

Thus, e.g. crystal growth kinetics can determine the conductivity type. In Fig. 7.21 the conductivity due to carbon background is shown for GaAs grown using MOVPE under various growth conditions. At high (low) arsine partial pressure incorporation of carbon on As-sites is less (more) probable, thus the conductivity is n-type (p-type). Also, growth on different surfaces can evoke different impurity incorporation, e.g. n-type on (001) GaAs and p-type on (311)A GaAs, since the latter is Ga-stabilized.

The charge density at an impurity nucleus can be investigated via the isomer shift as determined by Mössbauer spectroscopy [535, 536]. The incorporation of the isotope ^{119}Sn can be controlled in III-V compounds to be on cation or anion site as donor or acceptor, respectively. This is accomplished by introducing ^{119}In or ^{119}Sb on group-III and group-V site, respectively, both decaying into ^{119}Sn without leaving their lattice site. The isomer shifts of ^{119}Sn in various III-V compounds are shown in Fig. 7.22. In [536] it is concluded from these data that the tin donor is formed by a positive tin ion and the electron charge transfer to its neighboring (group-V) atoms is rather small. For tin as an acceptor, for the present conditions an ionized, i.e. negatively charged acceptor, the isomer shift follows closely the trend from substitution in group-IV semiconductors. Therefore four electrons form the tetrahedral bond, while the extra electron is located rather at the (positively charged) group-III next neighbors and not in the impurity cell. The difference to the point charge Coulomb distribution is called central-cell correction.

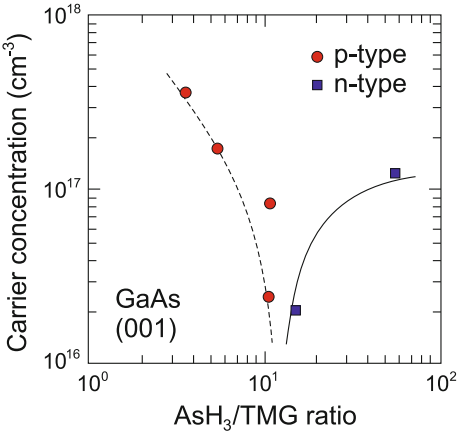

Fig. 7.21 Background doping of GaAs due to carbon in MOVPE for different ratios of the partial pressures of AsH_3 and TMG (trimethylgallium). The conductivity type (*blue squares* n-type, *red circles* p-type) depends on the incorporation of C from CH_3 radicals on Ga- or As-site. *Lines* are guides to the eye. Experimental data from [534]

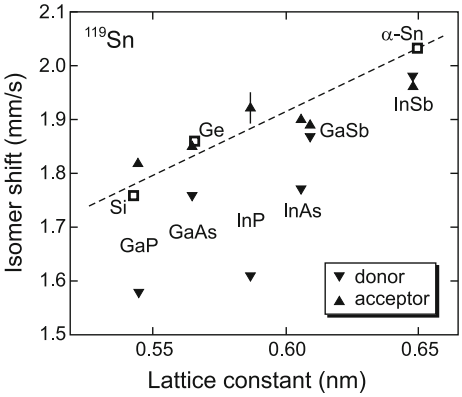

Fig. 7.22 Isomer shift (relative to $CaSnO_3$) of ^{119}Sn in various group-IV and III-V compound semiconductors as labeled. *Dashed line* is trend from isoelectronic substitution. Experimental data from [536]

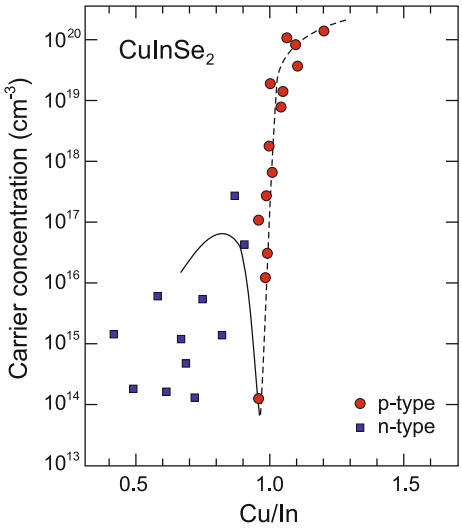

Fig. 7.23 Carrier concentration and conductivity type (*blue squares* n-type, *red circles* p-type) as a function of stoichiometry for $CuInSe_2$ thin films. *Lines* are guides to the eye. Experimental data from [537]

Deviation from the ideal stoichiometry introduces point defects that can be electrically active and change conductivity type and carrier concentration. In the case of $CuInSe_2$, excess Cu could go on interstitial positions or promote selenium vacancies, both leading to n-type behavior. This material is particularly sensitive to deviations from ideal stoichiometry for both Cu/In ratio (Fig. 7.23) and Se deficiency [537].

7.5.6 Autodoping

If intrinsic defects such as vacancies or interstitials, possibly as a result of non-stoichiometry, or anti-site defects cause electronic levels relevant for conductivity one speaks of autodoping. An example is the role of A-B antisites in AB_2O_4 spinels

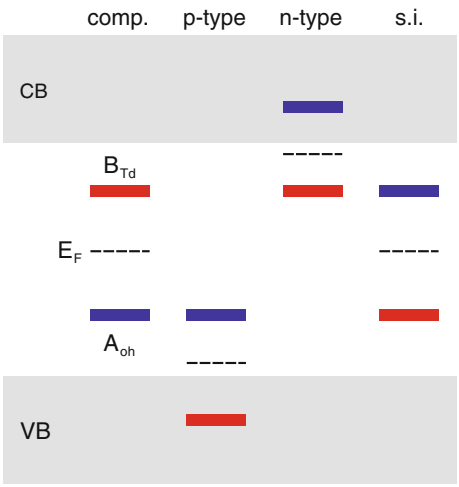

Fig. 7.24 Schematic position of electronic levels of A_B (*blue*, 0/− transition level) and B_A (*red*, +/0 transition level) defects in AB_2O_4 spinels and resulting material properties (compensated, n- or p-type or semi-insulating). After [538]

(Sect. 3.4.7). In the perfect crystal the A (B) atoms occupy tetraeder (octaeder) places. Typical charges are A^{2+} and B^{3+}. Thus (without charge transfer) the A atom on octaeder site (A_{Oh}) acts like a donor and the B atom on a tetraeder site (B_{Td}) as an acceptor. Such defects have been classified in [538] as being able to create compensated, semi-insulating, n-type or p-type material depending on the defect formation energies and the position of the electronic levels of the A_B and B_A defects in the band gap (Fig. 7.24). An example for a p-type spinel oxide is $ZnCo_2O_4$ [539].

7.5.7 *High Doping*

For low doping concentrations, the impurity atoms can be considered to be decoupled. At low temperature, only hopping from one impurity to the next is possible due to thermal emission or tunneling and the semiconductor becomes an insulator.

With increasing concentration, the distance between impurities decreases and their wavefunctions can overlap. Then, an impurity band develops (Fig. 7.25). A periodic arrangement of impurity atoms would result in well-defined band edges as found in the Kronig–Penney model. Since the impurity atoms are randomly distributed, the band edges exhibit tails. For high doping, the impurity band overlaps with the conduction band. In the case of compensation, the impurity band is not completely filled and contains (a new type of) holes. In this case, conduction can take place within the impurity band even at low temperature, making the semiconductor a metal. This metal–insulator transition has been discussed by Mott [540]. Examples for highly doped semiconductors are transparent conductive oxides (Chap. 20), the contact layer for an ohmic contact (Sect. 21.2.6) or the active

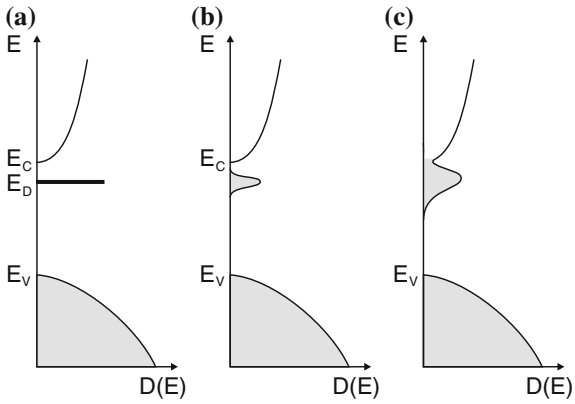

Fig. 7.25 Principle of the formation of a (donor) impurity band. (**a**) Small doping concentration and sharply defined impurity state at E_D, (**b**) increasing doping and development of an impurity band that (**c**) widens further and eventually overlaps with the conduction band for high impurity concentration. The *shaded areas* indicate populated states at $T = 0\,\mathrm{K}$

layers in a tunneling diode (Sect. 21.5.9). The physics, properties and preparation of highly doped semiconductors are treated in detail in [541].

The formation of the impurity band leads to a reduction of the impurity ionization energy as known from (7.19). Typical results are shown in Fig. 7.26a for n-type Ge [515] and Fig. 7.26b for p-type ZnTe [542]. At the critical doping concentration of $N_c = 1.5 \times 10^{17}$, the activation energy for the carrier concentration disappears. Similar effects have been observed for Si [543] and GaAs [544]. The freeze-out of the carrier concentration (see Fig. 7.9) disappears as shown in Fig. 7.27. Critical doping concentrations are listed in Table 7.6. The decrease of the ionization energy E^b (donor or acceptor) follows the dependence [515, 543]

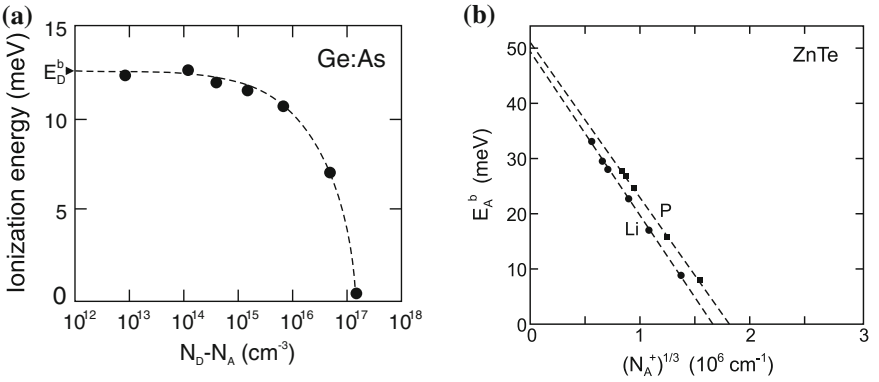

Fig. 7.26 (**a**) Donor ionization energy in n-type Ge for various doping concentrations. *Dashed line* is a guide to the eye. The arrow labeled E_D^b denotes the low-concentration limit (cf. Table 7.2). Experimental data from [515]. (**b**) Acceptor ionization energy for ZnTe:Li and ZnTe:P as a function of the third root of the ionized acceptor concentration. Data from [542]

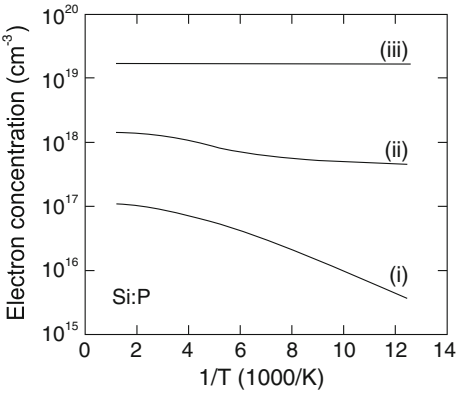

Fig. 7.27 Electron concentration versus inverse temperature for Si:P for three different doping concentrations ((i): $1.2 \times 10^{17}\,\mathrm{cm}^{-3}$, (ii): $1.25 \times 10^{18}\,\mathrm{cm}^{-3}$, (iii): $1.8 \times 10^{19}\,\mathrm{cm}^{-3}$). Experimental data from [543]

Table 7.6 Critical doping concentration for various semiconductors (at room temperature)

Material	Type	$N_c\,(\mathrm{cm}^{-3})$	Ref.
C:B	p	2×10^{20}	[518]
Ge:As	n	1.5×10^{17}	[515]
Si:P	n	1.3×10^{18}	[543]
Si:B	p	6.2×10^{18}	[543]
GaAs	n	1.0×10^{16}	[544]
GaP:Si	n	6×10^{19}	[546]
GaP:Zn	p	2×10^{19}	[547]
GaN:Si	n	2×10^{18}	[548]
GaN:Mg	p	4×10^{20}	[519]
$Al_{0.23}Ga_{0.77}N$:Si	n	3.5×10^{18}	[549]
ZnTe:Li	p	4×10^{18}	[542]
ZnTe:P	p	6×10^{18}	[542]
ZnO:Al	n	8×10^{18}	[550]

$$E^{\mathrm{b}} = E_0^{\mathrm{b}} - \alpha\, N_{\mathrm{i}}^{1/3} = E_0^{\mathrm{b}} \left[1 - \left(\frac{N_{\mathrm{i}}}{N_{\mathrm{c}}} \right)^{1/3} \right], \tag{7.53}$$

where N_{i} is the concentration of ionized dopants. A refined theory, considering screening, shift and tails of the conduction band and most importantly broadening of the donor level has been presented in [545].

The critical density can be estimated from the Mott criterion when the distance of the impurities becomes comparable to their Bohr radius (7.20)

$$2a_{\mathrm{D}} = \frac{3}{2\pi}\, N_{\mathrm{c}}^{1/3}. \tag{7.54}$$

Table 7.7 Maximum electrically active doping concentration for GaAs

Material	Type	N_c (cm^{-3})	Ref.
GaAs:Te	n	2.6×10^{19}	[554]
GaAs:Si	n	1.8×10^{19}	[555]
GaAs:C	p	1.5×10^{21}	[556]
GaAs:Be	p	2×10^{20}	[557]

The pre-factor $3/(2\pi)$ stems from the random distribution of impurities and disappears for a periodic arrangement. The Mott criterion is (rewriting (7.54))

$$a_D N_c^{1/3} \approx 0.24. \tag{7.55}$$

For GaAs with $a_D = 10.3\,nm$, the criterion yields $N_c = 1.2 \times 10^{16}\,cm^{-3}$, in agreement with experiment.

The achievable maximum concentration of electrically active dopants is limited by the concentration dependence of the diffusion coefficient, Coulomb repulsion, autocompensation and the solubility limit [501]. In Table 7.7 the maximum carrier concentrations for GaAs with various dopants are listed.

As an example we show the Ga-doping of epitaxial ZnO layers on sapphire in Fig. 7.28. Under slightly Zn-rich (O-polar) conditions the growth mode is two-dimensional and the carrier concentration increases linearly with the Ga concentration, $n \approx c_{Ga}$, up to high values in the $10^{20}\,cm^{-3}$ range [551]. For O-rich (Zn-polar) conditions the growth mode changes to three-dimensional growth and the activation ratio of Ga donors becomes low [552]. Above a gallium content of 2 %, the octahedral coordination of gallium and thus the partial segregation into a parasitic $ZnGa_2O_4$ spinel phase is observed for [Ga] = 4 % [553].

The random distribution of dopants essentially makes a doped semiconductor a disordered system. The physics of electronic states in disordered systems has been reviewed in [558]. A metal–insulator transition is observed at a certain value of

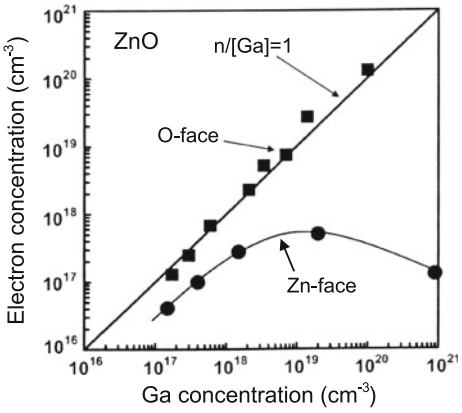

Fig. 7.28 Electron concentration as a function of gallium concentration in MBE grown ZnO:Ga on sapphire for the two different polarities. Adapted from [551, 552]

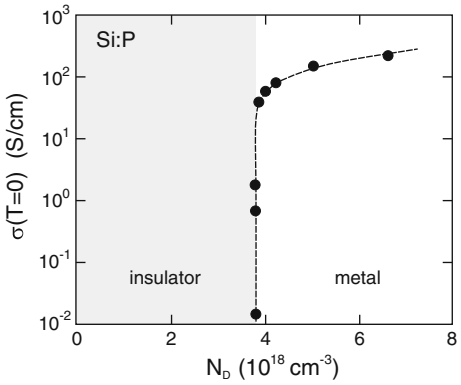

Fig. 7.29 Zero temperature conductivity of Si:P for various (donor) doping concentrations. Experimental data (*symbols*) and guide to the eye (*dashed line*). Adapted from [559]

doping ($N_P = 3.8 \times 10^{18}\,\mathrm{cm}^{-3}$), as shown in Fig. 7.29 for Si:P [559]. For a certain value of disorder all states become localized (Anderson localization [560], cmp. Sect. 8.8).

7.6 Quasi-Fermi Levels

The carrier concentrations were given by (7.6) and (7.7). So far, we have only considered semiconductors in thermodynamic equilibrium for which $np = n_i^2$. In a non-equilibrium situation, e.g. for external excitation or carrier injection in a diode, the electron and hole densities can each take arbitrary values, in principle. In particular, np will no longer be equal to n_i^2 and there is no Fermi level constant throughout the structure. In this case, however, quasi-Fermi levels F_n and F_p for electrons and holes, respectively, are defined via

$$n(\mathbf{r}) = N_C\, F_{1/2}\left(\frac{F_n(\mathbf{r}) - E_C}{kT}\right) \tag{7.56a}$$

$$p(\mathbf{r}) = N_V\, F_{1/2}\left(-\frac{F_p(\mathbf{r}) - E_V}{kT}\right). \tag{7.56b}$$

A quasi-Fermi level is sometimes called imref[4] and can also be denoted as E_{F_n} or E_{F_p}. We emphasize that the quasi-Fermi levels are *only* a means to describe the local carrier density in a logarithmical way. The quasi-Fermi levels can be obtained from the density via

[4] W. Shockley had asked E. Fermi for permission to use his name reversed. Fermi was not too enthusiastic but granted permission.

$$F_\mathrm{n} = E_\mathrm{C} + kT \ln\left(\frac{n}{N_\mathrm{C}}\right) \tag{7.57a}$$

$$F_\mathrm{p} = E_\mathrm{V} - kT \ln\left(\frac{p}{N_\mathrm{V}}\right). \tag{7.57b}$$

The quasi-Fermi levels do not imply that the carrier distribution is actually a Fermi distribution. This is generally no longer the case in thermodynamical nonequilibrium. However, in 'well-behaved' cases the carrier distribution in nonequilibrium can be approximated locally as a Fermi distribution using a local quasi-Fermi level and a local temperature, i.e.

$$f_\mathrm{e}(\mathbf{r}, E) \cong \frac{1}{\exp\left(\frac{E - F_\mathrm{n}(\mathbf{r})}{kT(\mathbf{r})}\right) + 1}. \tag{7.58}$$

Using the quasi-Fermi levels, np is given by

$$n(\mathbf{r})\, p(\mathbf{r}) = n_\mathrm{i}^2 \exp\left(\frac{F_\mathrm{n}(\mathbf{r}) - F_\mathrm{p}(\mathbf{r})}{kT}\right). \tag{7.59}$$

We note that for an inhomogeneous semiconductor or a heterostructure (cf. Chap. 12), n_i may also depend on the spatial position. In the case of thermodynamic equilibrium the difference of the quasi-Fermi levels is zero, i.e. $F_\mathrm{n} - F_\mathrm{p} = 0$ and $F_\mathrm{n} = F_\mathrm{p} = E_\mathrm{F}$.

7.7 Deep Levels

For deep levels the short-range part of the potential determines the energy level. The long-range Coulomb part will only lead to a correction. The term 'deep level' implies that the level is within the band gap and far from the band edges. However, some deep levels (in the sense of the potential being determined by the ion core) have energy levels close to the band edges or even within a band. Details can be found in [233, 561–564].

The wavefunction is strongly localized. Thus, it cannot be composed of Bloch functions, as has been done for the shallow levels for the effective-mass impurity. The localization in $\mathbf{r}$ space leads to a delocalization in $\mathbf{k}$ space. Examples are Si:S, Si:Cu or InP:Fe, GaP:N, ZnTe:O. Deep levels can also be due to intrinsic defects such as vacancies or antisite defects.

Due to the larger distance to the band edges, deep levels are not efficient at providing free electrons or holes. Quite the opposite, they rather capture free carriers and thus lead to a reduction of conductivity. Centers that can capture electrons *and* holes lead to nonradiative recombination of electrons through the deep level into the valence band (see also Chap. 10). This can be useful for the fabrication of semi-insulating

layers with low carrier concentration and fast time response of, e.g. switches and photodetectors.

While the electronic properties of deep levels can be readily characterized, the microscopic origin is not immediately apparent. Next to theoretical modeling of defects and correlation with experimental results, paramagnetic hyperfine interactions have proven useful to identify the microscopic nature of various defects [565].

7.7.1 Charge States

The deep level can have different charge states depending on the occupancy with electrons. The energy position within the gap varies with the charge state due to the Coulomb interaction. Also, the lattice relaxation around the defect depends on the charge state and modifies the energy level.

The localized charge q_d at the defect is the integral over the change $\Delta\rho$ of the charge density compared to the perfect crystal over a sufficiently large volume V_∞ around the defect

$$q_\mathrm{d} = \int_{V_\infty} \Delta\rho(\mathbf{r})\, \mathrm{d}^3\mathbf{r} = \frac{n\,e}{\epsilon_\mathrm{r}}. \tag{7.60}$$

In semiconductors, the charge $q_\mathrm{d}\epsilon_\mathrm{r}$ is an integer multiple of the elementary charge. The defect is said to be in the nth charge state. Each charge state has a certain stable atomic configuration $\mathbf{R}_n$. Each charge state has a ground state and excited states that can each have different stable atomic configurations.

Now, we discuss how the concentration of the various charge states depends on the position of the Fermi level. The overall constraint of global charge neutrality determines the chemical potential of the electron, i.e. the Fermi level in Fermi–Dirac statistics. We use the approximation that the concentration of defects is so small that the mutual interaction of defects becomes negligible.

As an example, we treat the possible reaction $V^0 \rightleftharpoons V^+ + e^-$, where V^0 denotes a neutral vacancy and V^+ is a positively charged vacancy, created by the ionization of an electron from the vacancy into the conduction band. The free energy G depends on the numbers n_0 of neutral and n_+ of positively charged vacancies. The minimum condition is met by

$$\mathrm{d}G = \frac{\partial G}{\partial n_0}\,\mathrm{d}n_0 + \frac{\partial G}{\partial n_+}\,\mathrm{d}n_+ = 0. \tag{7.61}$$

The neutrality constraint is $\mathrm{d}n_0 + \mathrm{d}n_+ = 0$ and therefore the minimum condition reads

$$\frac{\partial G}{\partial n_0} = \frac{\partial G}{\partial n_+}. \tag{7.62}$$

For noninteracting defects and using (4.9) we write

$$\frac{\partial G}{\partial n_0} = G^{\mathrm{f}}(V^0) + kT \ln\left(\frac{n_0}{N_0}\right) \tag{7.63a}$$

$$\frac{\partial G}{\partial n_+} = \frac{\partial G(V^+)}{\partial n_+} + \frac{\partial G(e^-)}{\partial n_+} = G^{\mathrm{f}}_{V^+} + kT \ln\left(\frac{n_+}{N_+}\right) + \mu_{\mathrm{e}^-}, \tag{7.63b}$$

where $N_0 = NZ_0$ and $N_+ = NZ_+$ are the number of available sites, given by the number N of atomic sites and including possible internal degeneracies Z_0 and Z_+, respectively. Degeneracy factors of deep levels are not a simple subject [522] and, e.g. the degeneracy factors of Au donor and acceptor levels in Si are under discussion [566–568]. G_{f} denotes the free enthalpy of formation of the respective defect, as in (4.3). We have written the free enthalpy of the separated pair V^+ and e^- as the sum $G(V^+) + G(e^-)$. $\mu_{\mathrm{e}^-} = \partial G(e^-)/\partial n_+$ is (by definition) the chemical potential of the electron, i.e. the Fermi energy E_{F} of Fermi–Dirac statistics.[5] From (7.63a, b) we find for the ratio of the concentrations of defects $c_0 = n_0/N$ and $c_+ = n_+/N$

$$\frac{c_0}{c_+} = \frac{Z_+}{Z_0} \exp\left(-\frac{G^{\mathrm{f}}_{V^+} - G^{\mathrm{f}}_{V^0} + E_{\mathrm{F}}}{kT}\right) = \frac{Z_+}{Z_0} \exp\left(\frac{E_{\mathrm{t}}(V^0) - E_{\mathrm{F}}}{kT}\right), \tag{7.64}$$

where the trap level energy (for the particular charge transition), $E_{\mathrm{t}}(V^0) = G^{\mathrm{f}}_{V^0} - G^{\mathrm{f}}_{V^+}$, is the free enthalpy of ionization of V^0. We note that c_0 can be obtained from (4.9) and E_{F} is determined by the charge-neutrality condition.

As example experimental data on the charge transition $\mathrm{Fe}^0 \rightleftharpoons \mathrm{Fe}^+ + \mathrm{e}^-$ of interstitial iron (in tetrahedral position, Fig. 7.30a, cmp. Fig. 3.18) in silicon is shown. The concentration of Fe^0 is tracked via the EPR signal from the neutral $S = 1$ state[6] with g-factor $g = 2.07$ [569]. For n-type samples the iron is in neutral state and the maximum EPR signal is found. For strongly p-type samples, the Fermi energy is below the trap level and all iron is in Fe^+ state, yielding no EPR signal at the given g-factor. From the investigation of various silicon samples with different doping levels and consequently different position of the Fermi level, the trap (deep donor) energy is found to be $E_{\mathrm{V}} + 0.375\,\mathrm{eV}$ as indicated in Fig. 7.30b.

[5] The chemical potential in a one-component system is $\mu = \partial G/\partial n = G/n$. In a multicomponent system it is, for the ith component, $\mu_i = \partial G/\partial n_i \neq G/n_i$.

[6] The electron configuration is $3\mathrm{d}^8$ with two paramagnetic electrons. Under uniaxial stress along [100] the EPR line splits into a doublet. [569] Further details can be found in [570].

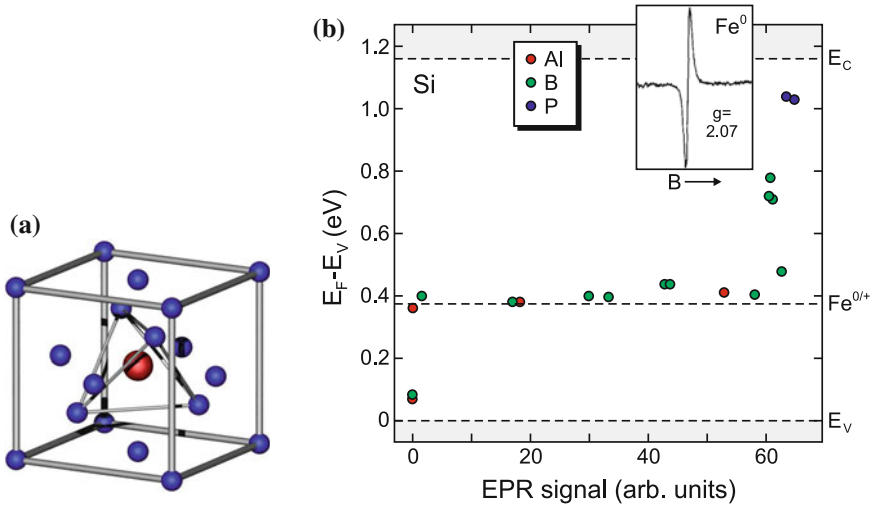

Fig. 7.30 (**a**) Silicon cubic unit cell with an interstitial iron atom (*red*) at tetrahedral site. (**b**) EPR intensity (at $T = 95\,\mathrm{K}$ from interstitial iron in neutral state, Fe^0 with $S = 1$) versus Fermi level position for iron-doped silicon with varying Fermi level due to different amounts of shallow impurity levels from to Al, B and P as labeled. The *shaded areas* indicate the valence and conduction band. The *dashed line* at $E_t = E_V + 0.375\,\mathrm{eV}$ indicates the trap level. The *inset* shows a typical EPR spectrum of Fe^0. Adapted from [571], *inset* adapted from [572]

7.7.2 Double Donors

An impurity that has two extra electrons available after bonding in the matrix may give rise to a double donor. Typical examples are substitutional chalcogenide atoms (S, Se or Te) in silicon [573] and germanium [574], interstitial impurities such as Mg_i in Si [575], or group-V atoms on a group-III site in a III–V compound (antisite defect), such as P_{Ga} in GaP [576] or As_{Ga} in GaAs [577].

The double donor is electronically similar to a helium atom. Due to the repulsive Coulomb interaction of the two electrons on the neutral double donor, the (single) ionization energy E_1 (also often labeled $E(0, 1)$ or $E(0, +)$) of D^0 is smaller than that of D^+ (E_2, also labeled $E(1, 2)$ or $E(+, ++)$). For He and He^+ the ratio of ionization energies is 0.45; for chalcogenides in Si and Ge similar ratios have mostly been found (Table 7.8).

The carrier statistics and the degeneracy factors for a double donor have been discussed in [510, 578]. Typically, the degeneracy factor for the ionization of the double donor $D^0 \rightarrow D^+$ is $\hat{g}_D = g_2/g_1 = 1/2$ and for the ionization $D^+ \rightarrow D^{++}$ is $\hat{g}_D = g_1/g_0 = 2/1 = 2$.

For the probabilities to find a neutral, single and double ionized donor we find following the treatment in [578]

Table 7.8 Binding energies (to conduction band) of double donor chalcogenide impurities in Si and Ge

Host	State	S	Se	Te
Si	D^0	318	307	199
	D^+	612	589	411
Ge	D^0	280	268	93
	D^+	590	512	330

All energies in meV, data from [573, 574]

$$d^0 = \frac{N_D^0}{N_D} = \frac{\exp\frac{2E_F}{kT}}{\exp\frac{E_1+E_2}{kT} + \exp\frac{2E_F}{kT} + 2\exp\frac{E_1+E_F}{kT}} \tag{7.65a}$$

$$d^+ = \frac{N_D^+}{N_D} = \frac{\exp\frac{E_1+E_2}{kT}}{\exp\frac{E_1+E_2}{kT} + \exp\frac{2E_F}{kT} + 2\exp\frac{E_1+E_F}{kT}} \tag{7.65b}$$

$$d^{++} = \frac{N_D^{++}}{N_D} = \frac{2\exp\frac{E_1+E_F}{kT}}{\exp\frac{E_1+E_2}{kT} + \exp\frac{2E_F}{kT} + 2\exp\frac{E_1+E_F}{kT}} \tag{7.65c}$$

The probabilities are depicted in Fig. 7.31a. The maximum of d^+ is at the energy $(E_1 + E_2)/2$. Its value is

$$d^+\left(\frac{E_1+E_2}{2}\right) = \frac{1}{1+\exp\left(-\frac{E_1-E_2}{2kT}\right)} \tag{7.66}$$

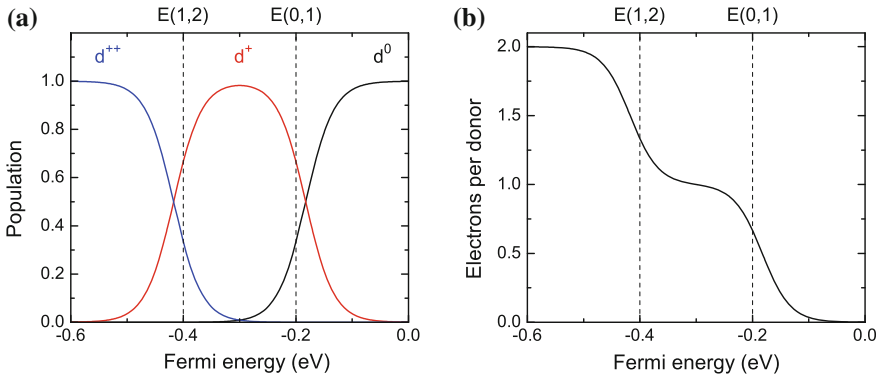

Fig. 7.31 (**a**) Population of states of a double donor (neutral: *black*, single ionized: *red*, double ionized: *blue*) according to (7.65a–c) as a function of the Fermi level. The ionization energies have been chosen as $E_1 = -0.2\,\text{eV}$ and $E_2 = -0.4\,\text{eV}$ and are indicated by *dashed lines* ($kT = 25\,\text{meV}$); these energies are similar to Si:Te (cmp. Table 7.8). The conduction band edge is taken as zero energy. (**b**) depicts the according number of electrons $\tilde{n}$ ionized from the donor.

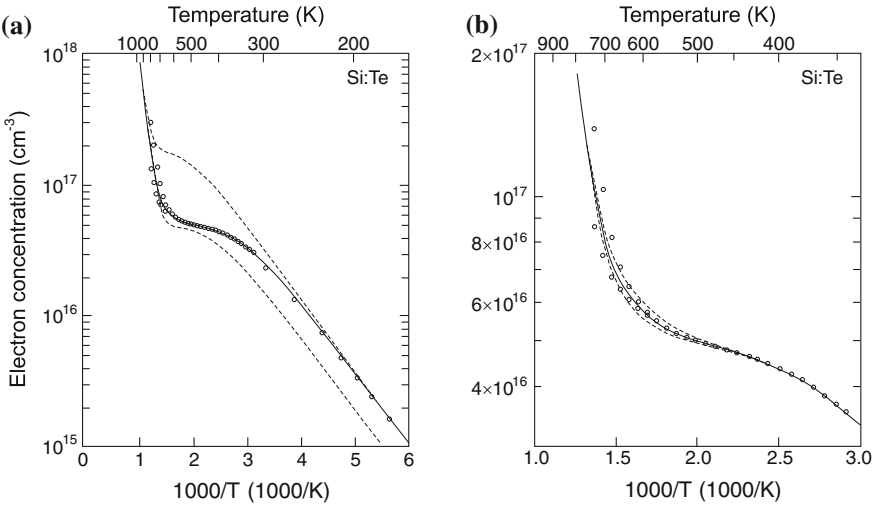

Fig. 7.32 Temperature dependent electron concentration (from Hall data) for Si:Te. (**a**) Experimental data and fit with double donor model using $N_{\mathrm{Te}} = 5 \times 10^{16}\,\mathrm{cm}^{-3}$, $E_1 = 200\,\mathrm{meV}$ and $E_2 = 440\,\mathrm{meV}$ (*solid line*). Single donor models would fail ($N_{\mathrm{Te}} = 5 \times 10^{16}\,\mathrm{cm}^{-3}$ and $N_{\mathrm{Te}} = 2 \times 10^{17}\,\mathrm{cm}^{-3}$, *dashed lines*). (**b**) Second ionization step in more detail with fits using different values for E_2; the *solid line* is for $E_2 = 440\,\mathrm{meV}$, the other *dashed lines* for $E_2 = 420$ and $460\,\mathrm{meV}$. Adapted from [579]

and reaches a value close to one for $(E_1 - E_2)/kT \gg 1$. In Fig. 7.31b the number of electrons per donor $\tilde{n} = (N_{\mathrm{D}}^{+} + 2\,N_{\mathrm{D}}^{++})/N_{\mathrm{D}}$ is shown as a function of the Fermi level; at $(E_1 + E_2)/2$, exactly $\tilde{n} = 1$. In Fig. 7.32 the temperature dependent electron concentration in Si:Te is depicted. Up to 570 K the single ionization is visible (other shallow impurities present in the sample in lower concentrations $< 10^{14}\,\mathrm{cm}^{-3}$ play no role). From the fit $E_1 = 200 \pm 2.7\,\mathrm{meV}$ is determined [579]. Single donor models would fail. The second ionization step is somewhat masked by the onset of intrinsic conduction. According to (7.15), the slope of n_i is $E_{\mathrm{g}}/2 \approx 500\,\mathrm{meV}$ which is similar to $E_2 \approx 440\,\mathrm{meV}$.

7.7.3 Double Acceptors

In analogy to double donor defects, double acceptors can introduce up to two holes into the valence band. A typical example is Zn in silicon [580], exhibiting its 'normal' acceptor level ($\mathrm{Zn}^0/\mathrm{Zn}^-$) at $E_{\mathrm{V}} + 0.31\,\mathrm{eV}$. In moderately n-doped silicon another level ($\mathrm{Zn}^-/\mathrm{Zn}^{2-}$) is observed at $E_{\mathrm{C}} - 0.55\,\mathrm{eV}$, when the n-doping is sufficient to partially compensate the Zn and supply one electron for each Zn atom but not two ($2N_{\mathrm{Zn}} > N_{\mathrm{D}} > N_{\mathrm{Zn}}$). A similar situation has been observed for Zn in germanium, exhibiting the levels $E_{\mathrm{V}} + 0.03\,\mathrm{eV}$ and $E_{\mathrm{V}} + 0.09\,\mathrm{eV}$ [581]. In Fig. 7.33 three

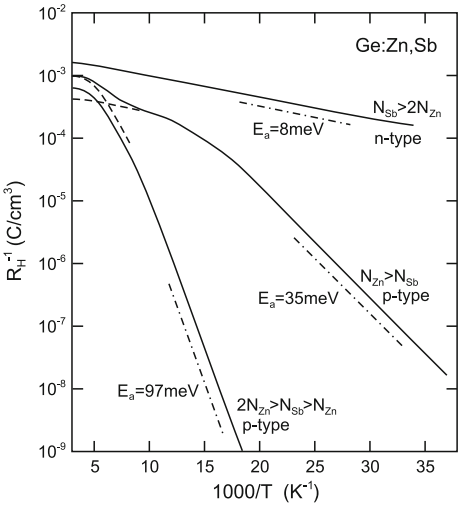

Fig. 7.33 Inverse (absolute) Hall coefficient (cmp. Sect. 13.2.1) R_H^{-1}, i.e. charge concentration, for three Ge:Zn samples with different degree of compensation with Sb donors as labeled. The *dash-dotted lines* indicate typical slopes. The *dashed lines* sketch the $Zn^0 \rightarrow Zn^-$ and the $Zn^- \rightarrow Zn^{--}$ processes. Adapted from [581]

different Ge:Zn samples are compared. If the additional Sb donor concentration ($N_D \approx 3.4 \times 10^{16}\,cm^{-3}$) is larger than $2N_{Zn}$ ($N_{Zn} \approx 1.2 \times 10^{16}\,cm^{-3}$), the sample is n-type (upper curve). The slope is similar to the Ge:Sb donor binding energy (Table 7.2). If compensation with donors is weak ($N_{Zn} > N_D$, middle curve) first the shallow donor level with 0.03 eV activation energy is activated and subsequently the deeper one with 0.09 eV activation energy, creating p-conduction with a saturated hole density of $p \approx 2N_A - N_D > N_{Zn}$ (negative Hall coefficient). The two individual activation processes are sketched as dashed lines in Fig. 7.33. If the Sb concentration is larger than N_{Zn} but smaller than $2N_{Zn}$, the shallow acceptor level is filled with electrons, leaving still the only partially filled deeper acceptor level available for ionization (lower curve). In this case the sample is still p-type, but the saturation hole density is $p \approx 2N_A - N_D < N_{Zn}$. The degeneracy factors for Zn in Si and Ge have been discussed in [522].

7.7.4 Jahn–Teller Effect

The lattice relaxation can reduce the symmetry of the defect. Many defects, such as a vacancy, a tetrahedral interstitial or an impurity, occupy initially tetrahedral sites in the zincblende structure. The lattice relaxation reduces the symmetry, e.g. to tetragonal or trigonal, and therefore causes initially degenerate levels to split. Such splitting is called the static Jahn–Teller effect [561, 582]. The energy change in terms of the atomic displacement Q can be denoted (using perturbation theory for the simplest, nondegenerate case) as $-I\,Q^2$ ($I > 0$). Including the elastic contribution

with a force constant C, the energy of a configuration Q is

$$E = -I\,Q + \frac{1}{2}\,C\,Q^2. \tag{7.67}$$

The stable configuration $Q_{\min}$, for which the energy is minimal ($E_{\min}$), is therefore given by

$$Q_{\min} = \frac{I}{C} \tag{7.68a}$$

$$E_{\min} = -\frac{I^2}{2C}. \tag{7.68b}$$

Several equivalent lattice relaxations may exist, e.g. a 3-fold minimum for remaining C_{3v} symmetry. The energy barrier between them has a finite height. Therefore, e.g. at sufficient temperature, the defect can switch between different configurations and eventually again becomes isotropic (dynamic Jahn–Teller effect). The experimental observation depends on the relation between the characteristic time of the experiment and the reorientation time constant of the defect.

7.7.5 *Negative-U Center*

We explain the principle of a so-called negative-U center [583] for the Si vacancy [584] (cf. Fig. 4.2). It was first proposed by Anderson to explain the properties of amorphous chalcogenide glasses [585]. Many defects in semiconductors exhibit negative-U behavior, e.g. also the boron interstitial in Si [584, 586]. Coulomb energy and the Jahn–Teller effect compete for the position of the occupancy level for different charge states. U refers to the additional energy upon charging of the defect with an additional electron. The Coulomb repulsion of electrons leads to an *increase* of the energy, i.e. positive U, which has been calculated to be 0.25 eV for the Si vacancy [587] for all charge states. The occupation level (cf. Sect. 4.2.2) $E_0(1, 2)$ (the index 0 indicates effects only due to many-electron Coulomb interaction), separating the domination of V^{++} and V^{+} (Fig. 7.34) is 0.32 eV above the valance-band edge. Therefore, the occupation level $E_0(0, 1)$ is expected to lie at about 0.57 eV about E_V.

The Jahn–Teller effect may lead to a splitting of the otherwise 4-fold degenerate states of the vacancy. A detailed experimental study using hyperfine interactions can be found in [588]. The schematic level diagram for the Jahn–Teller splitting is shown in Fig. 7.35. The V^{++} state (A_1 is always populated with two electrons) is resonant with the valence band. The T_2 state lies in the band gap. When the Jahn–Teller effect (now on the T_2 state) is included, the energies of the different charge states depend on the configuration coordinate (a mostly tetragonal distortion in the case of the Si vacancy).

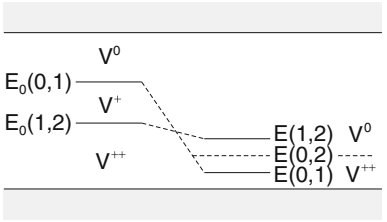

Fig. 7.34 Charge states of the vacancy in silicon. *Left* level scheme without lattice relaxation, *right* level scheme including the Jahn–Teller effect. For a Fermi level below (above) $E(0, 2)$ the charge state V^{++} (V^0) is dominant

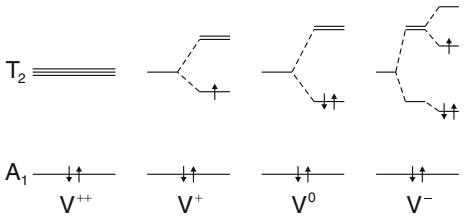

Fig. 7.35 Jahn–Teller splitting for different charge states of the vacancy. A_1 and T_2 refer to irreducible representations of the T_d point symmetry group. A_1 is nondegenerate and therefore does not exhibit a Jahn–Teller effect. T_2 is triply degenerate. The arrows represent electrons and their spin orientation

$$E_{V^0} = E(0, Q) = E(0, Q = 0) - 2IQ + \frac{1}{2}CQ^2 \tag{7.69a}$$

$$E_{V^+} = E(1, Q) = E(1, Q = 0) - IQ + \frac{1}{2}CQ^2 \tag{7.69b}$$

$$E_{V^{++}} = E(2, Q) = E(2, Q = 0) + \frac{1}{2}CQ^2. \tag{7.69c}$$

For the $n = 2$ state the T_2 gap state is empty and thus no degeneracy and Jahn–Teller term arises. For $n = 1$ there is a linear Jahn–Teller term. The occupation with two electrons (V^0) causes an approximately twice as large Jahn–Teller splitting for the $n = 0$ state. The force constant is assumed to be independent of the charge state. The energies for the minimum configurations $Q^n_{\min}$ are therefore

$$E(0, Q^0_{\min}) = E(0, Q = 0) - 4\,\frac{I^2}{2C} \tag{7.70a}$$

$$E(1, Q^1_{\min}) = E(1, Q = 0) - \frac{I^2}{2C} \tag{7.70b}$$

$$E(2, Q^2_{\min}) = E(2, Q = 0). \tag{7.70c}$$

The Jahn–Teller energy $E_{\mathrm{JT}} = I^2/2C$ lowers the position of the occupancy levels E_0 calculated with Coulomb terms only. The occupancy levels *including* the Jahn–Teller contribution are therefore given as

$$E(1,2) = E_0(1,2) - E_{\mathrm{JT}} \tag{7.71a}$$

$$E(0,1) = E_0(0,1) - 3\,E_{\mathrm{JT}}. \tag{7.71b}$$

For the vacancy in silicon the Jahn–Teller energy E_{JT} is about 0.19 eV. Thus the $E(1,2)$ level is lowered from 0.32 to 0.13 eV. The E(0,1) occupancy level, however, is reduced from 0.57 to 0.05 eV [584, 589] (see Fig. 7.34). The occupancy level $E(0,2)$ is in the middle between $E(0,1)$ and $E(1,2)$ ($E(0,2) = (E(0,1) + E(1,2))/2$) and indicated in Fig. 7.36a. At this energy, $c(V^0) = c(V^{++})$ and the value of $c(V^+)$ is small ($\approx \exp\frac{E_1 - E_2}{2kT}$) since $E(0,1) < E(1,2)$ (cmp. (7.66)).

The relative concentrations of the three charge states are determined by (7.64) (degeneracy and entropy terms have been neglected)

$$\frac{c(V^{++})}{c(V^{+})} = \exp\left(\frac{E(1,2) - E_{\mathrm{F}}}{kT}\right) \tag{7.72a}$$

$$\frac{c(V^{+})}{c(V^{0})} = \exp\left(\frac{E(0,1) - E_{\mathrm{F}}}{kT}\right). \tag{7.72b}$$

They are depicted in Fig. 7.36a in a plot related to Fig. 7.31a. Therefore, V^{++} dominates if $E_{\mathrm{F}} < E(0,1)$ and V^0 dominates for $E_{\mathrm{F}} > E(1,2)$. In the intermediate range $E(0,1) < E_{\mathrm{F}} < E(1,2)$ we know from (7.72a, b) that V^+ is dominated by V^0 and V^{++}. However, at this point it is not clear whether V^{++} or V^0 dominates

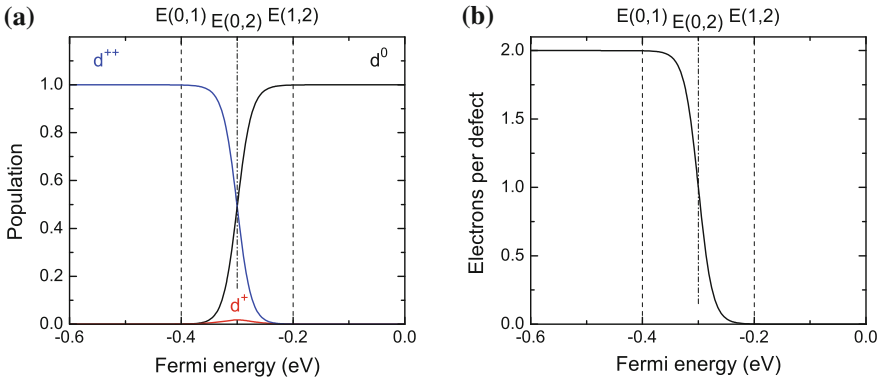

Fig. 7.36 (**a**) Population of states of a negative-U defect (neutral: *black*, single ionized: *red*, double ionized: *blue*) according to (7.65a–c) as a function of the Fermi level. The ionization energies have been chosen as $E_1 = -0.4$ eV and $E_2 = -0.2$ eV (cmp. Fig. 7.31) and are indicated by *dashed lines* ($kT = 25$ meV). The occupancy level $E(0,2)$ is indicated with a *dash-dotted line*. The conduction band edge is taken as zero energy. (**b**) depicts the according number of electrons ionized from the defect.

overall. The ratio of the concentrations of V^{++} and V^0 is given by

$$\frac{c(V^{++})}{c(V^0)} = \exp\left(\frac{E(1,2) + E(0,1) - 2E_F}{kT}\right) = e^2 \exp\left(\frac{E(0,2) - E_F}{kT}\right). \quad (7.73)$$

The occupancy level $E(0, 2)$ is thus again given as

$$E(0,2) = \frac{E(0,1) + E(1,2)}{2}, \quad (7.74)$$

and is shown in Fig. 7.34. V^{++} dominates if $E_F < E(0, 2)$ and V^0 dominates for $E_F > E(0, 2)$. V^+ is, for no position of the Fermi level, the dominating charge state of the Si vacancy. We note that for n-doped Si the V^- and V^{--} can also be populated. The population of the V^0 state with an extra electron introduces another Jahn–Teller splitting (Fig. 7.35) that has trigonal symmetry.

Generally, the Jahn–Teller effect can make the addition of an electron cause an effectively negative charging energy; in this case the center is termed a negative-U center. We note that the single vacancy in germanium is not a negative-U center due to smaller Jahn–Teller distortion and smaller electron-lattice coupling [590].

7.7.6 DX Center

The DX center is a deep level that was first investigated for n-doped (e.g. Si-doped) $Al_xGa_{1-x}As$. It dominates the transport properties of the alloy for $x > 0.22$. For smaller Al concentrations and GaAs the DX level lies in the conduction band. DX-type deep levels have also been found for other alloys and dopants, e.g. GaAsP:S.

It is experimentally found that the capture process of electrons into the DX center is thermally activated. The capture energy E_c depends on the AlAs mole fraction (Fig. 7.37). The (average) barrier for electron capture has a minimum of 0.21 eV for $x \approx 0.35$, near the crossover point between direct and indirect band gap

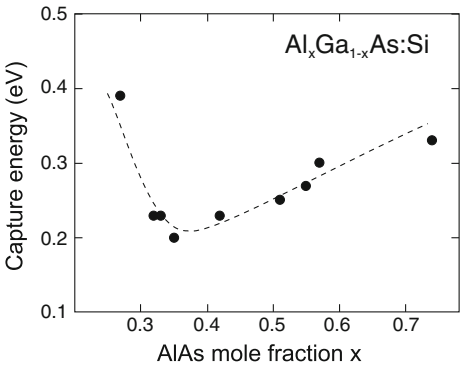

Fig. 7.37 Energy barrier for electron capture E_c at the Si-DX center in $Al_xGa_{1-x}As$ for various compositions. Experimental data from [591]

(cf. Fig. 6.19). For lower Al concentrations, the capture barrier increases to 0.4 eV for $x = 0.27$; for $x > 0.35$ the capture barrier increases to about 0.3 eV for x around 0.7 [591]. The barrier for thermally releasing carriers from the DX center has been determined to be about 0.43 eV, independent of the Al mole fraction [591].

Carriers can be removed from the DX center by optical absorption of photons with energy larger than about 1.2 eV. If carriers are removed by optical excitation at low temperatures the (re-)capture is so slow ($\sigma < 10^{-30}\,\mathrm{cm}^2$) that the carriers remain in the conduction band and cause persistent photoconductivity (PPC). The PPC is only reduced upon increasing the sample temperature. The concentration of the DX center is about the same as the net doping concentration.

The properties of the DX center are reviewed in [592, 593]. So far, no definite microscopic model of the DX center has been agreed on. Lang [594] proposed that the DX center involves a donor and an unknown defect (probably a vacancy). It probably involves large lattice relaxation as in the configuration coordinates model of Fig. 7.38 where the donor binding energy E_D^b with respect to the conduction-band minimum, the barrier for electron capture E_c, the barrier for electron emission E_e and the optical ionization energy E_o are labeled. The donor binding energy is measured with Hall effect (cf. Sect. 13.2.1) at temperatures sufficient to overcome the capture and emission barriers, the emission barrier is measured with deep level transient spectroscopy (DLTS). The capture barrier manifests itself in PPC experiments. We note that the DX center is related to the L-conduction band. For small Al mole fraction, the DX level is degenerate with the Γ-related conduction band (see Fig. 7.38b).

Theoretical models and experimental evidence hint at a vacancy-interstitial model for the Si-DX center [595]. The donor (Si) is displaced along the ⟨111⟩ direction from the Ga substitution site. The displacement is predicted to be 0.117 nm and the

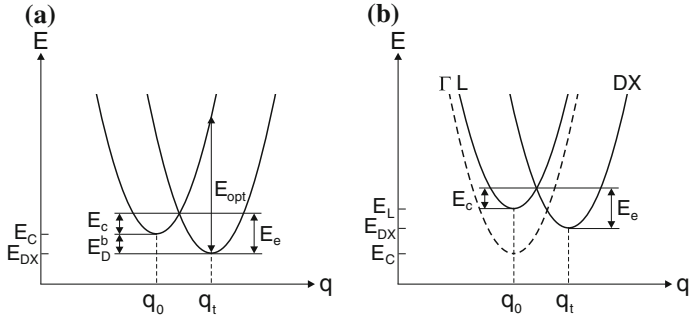

Fig. 7.38 (**a**) Schematic configuration coordinate diagram for the DX level with large lattice relaxation. q_0 is the configuration of the empty defect, q_t is the configuration of the filled defect. The donor binding energy E_D^b, the barrier for electron capture E_c, the barrier for electron emission E_e and the optical ionization energy E_o are labeled. E_C denotes the conduction-band edge. We note that in AlGaAs the DX level is associated with the L conduction band (see Fig. 6.19). (**b**) Schematic configuration coordinate diagram for the DX level in $Al_{0.14}Ga_{0.86}As$ with the DX level being degenerate with the (Γ-related) conduction band

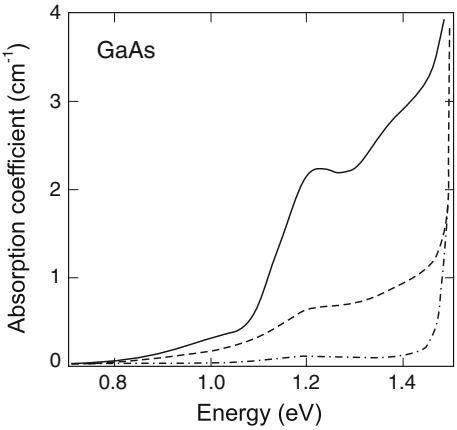

Fig. 7.39 Absorption spectrum of GaAs at low temperatures ($T = 10\,\mathrm{K}$) when cooled in the dark (*solid line*). The *dashed* (*dash-dotted*) *line* is the absorption after illuminating the sample for 1 min (10 min) with white light, leading to quenching of the EL2-related absorption. Adapted from [597]

distorted geometry can be viewed as a Ga vacancy and a Si interstitial. The charge state of the (filled) DX center is proposed to be a two-electron negative-U state.

7.7.7 EL2 Defect

The EL2 defect is a deep donor in GaAs. It is not related to impurities but occurs for intrinsic material, in particular grown under As-rich conditions. It has physical properties similar to the DX center. The bleaching of absorption due to EL2, i.e. the optical removal of electrons from the defect at low temperatures, is shown in Fig. 7.39. The microscopic model [596] describes the EL2 defect as an arsenic antisite defect, i.e. an arsenic atom on a Ga site, $\mathrm{As_{Ga}}$. In the charged state the arsenic atom is displaced from the lattice position and a complex of a Ga vacancy (symmetry T_{3d}) and an interstitial As (symmetry C_{3v}) with 0.14 nm displacement along $\langle 111 \rangle$ forms ($\mathrm{V_{Ga}}$-$\mathrm{As_i}$). The charged state is filled with two electrons.

7.7.8 Semi-insulating Semiconductors

Semiconductors with high resistivity (10^7–$10^9\,\Omega\,\mathrm{cm}$) are called semi-insulating ('s.i.' or 'si'). Semi-insulating substrates are needed for high-speed devices. The high resistivity should stem from a small free-carrier density at finite temperature and not from a small mobility due to poor crystal quality. For sufficiently wide band gap, the intrinsic carrier concentration is small and such pure material is semi-insulating, e.g. GaAs with $n_\mathrm{i} = 1.47 \times 10^6\,\mathrm{cm}^{-3}$ and $5.05 \times 10^8\,\Omega\,\mathrm{cm}$ [598]. Since shallow impurities are hard to avoid, another route is used technologically. Impurities that form deep levels are incorporated in the semiconductor in order to compensate free carriers. For

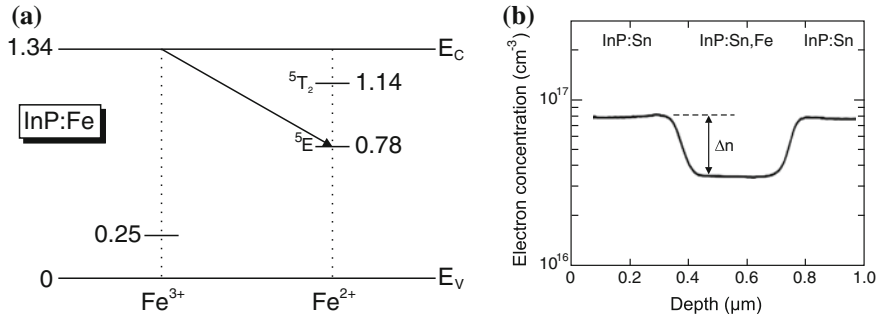

Fig. 7.40 (**a**) Schematic band diagram of InP with levels of Fe impurities in the 3+ and 2+ charge states at low temperature. All energies are given in eV. The *arrow* denotes capture of an electron (from the conduction band or a shallow donor) on the deep acceptor. Compare this figure also with Figs. 9.33 and 10.25. (**b**) Depth profile of electron concentration in an InP:Sn/InP:Sn,Fe/InP:Sn structure. The change $\Delta n \approx 4.5 \times 10^{16}\,\mathrm{cm}^{-3}$ of electron concentration is due to the compensation by Fe and corresponds to the chemical iron concentration determined by SIMS, $[\mathrm{Fe}] = 4.9 \times 10^{19}\,\mathrm{cm}^{-3}$. Part (**b**) adapted from [610]

example, a deep acceptor compensates all electrons if $N_A > N_D$. Since the acceptor is deep ($E_A^b \gg kT$), it does not release holes for reasonable temperatures. Examples of suitable impurities for compensation of electrons are Si:Au [599], GaAs:Cr [600] and InP:Fe [601]. A deep donor, e.g. InP:Cr [602], is necessary to compensate p-type conductivity.

Figure 7.40a shows the terms of Fe in InP [603, 604]. An overview of transition metals in III–V semiconductors can be found in [605]. The electron configuration of neutral Fe atoms is $3d^6 4s^2$ (cf. Table 16.2). The Fe is incorporated on the In site and thus has a Fe^{3+} state as a neutral acceptor (A^0). The Fe^{3+} state has the electron configuration $3d^5$. The arrow in Fig. 7.40a represents the capture of an electron from the conduction band or from a shallow donor. The charge state of the Fe becomes Fe^{2+} (charged acceptor, A^-) with the electron configuration $3d^6$. The cubic crystal field (T_d symmetry) splits this ^{5}D Fe state[7] into two terms [606] that exhibit further fine structure [604]. The large thermal activation energy of 0.64 eV found in the Hall effect on semi-insulating InP:Fe [601] corresponds to the energy separation of the ^{5}E level and the conduction band.

The maximum electron concentration that can be compensated in this way is limited by the solubility of Fe in InP [607], about $1 \times 10^{17}\,\mathrm{cm}^3$. Higher Fe incorporation leads to the formation of Fe (or FeP) precipitates and degrades the crystal quality. Only a fraction of the incorporated Fe may then be electrically active and contribute to the compensation. The maximum electrically active Fe concentration is found

[7]The notation is ^{2S+1}J (multiplicity), with S being the total spin and J being the total angular momentum.

to be $5–6\times10^{16}\,\mathrm{cm}^{-3}$ [608]. The compensation can be directly visualized via the depth profile of the electron concentration in a n-si-n structure (Fig. 7.40b). The poor thermal stability of Fe, i.e. high diffusion coefficient, has evoked proposals for more stable dopants such as InP:Ru [609].

7.7.9 Isoelectronic Impurities

Isoelectronic impurities, generally represent a deep level with a short range potential. The isoelectronic trap introduces a bound state for an electron or a hole. Once a carrier has been captured, the defect becomes charged. The other carrier type is then easily trapped, forming a bound exciton (Sect. 10.3.2). The theory of isoelectronic impurities is outlined in [611]. A detailed theoretical treatment of N in GaAs and GaP is given in [472].

In GaP:N, an electron is spatially localized on the N impurity. Most of the wave function is at the X-point. The nitrogen-bound electron level in GaP (A_1 symmetry) is close to the conduction band edge and within the band gap. Important for the energy position is the lattice relaxation, leading to an inward relaxation of the surrounding Ga atoms (Fig. 7.42). Due to the spatial localization of the wave function it is delocalized in k-space (Fig. 7.41a) and obtains a sizeable component at the Γ-point, facilitating zero-phonon absorption from the valence band. This effect is present only when the lattice relaxation around the impurity is considered; without relaxation the Γ-component is zero, with relaxation about 1 % [472]. The Γ-component of the wave-function is larger for localization at an isoelectronic impurity than at a shallow donor such as sulfur [612]. This way a large oscillator strength for optical transitions occurs

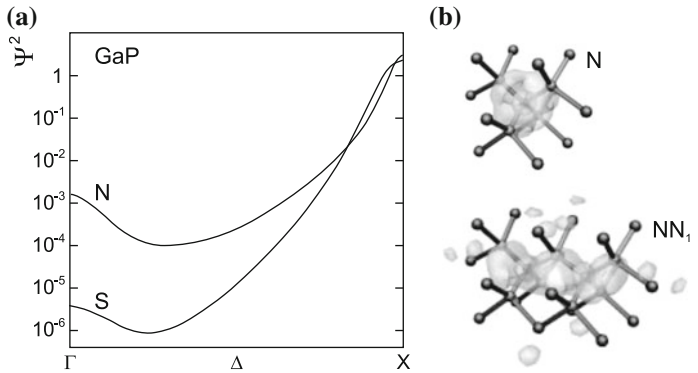

Fig. 7.41 (**a**) Model calculation of the wave-vector dependence of the probability density of an electron bound to a 10 meV deep isoelectronic trap (N) and to a 100 meV deep shallow donor (S) in GaP. Adapted from [612]. (**b**) Wavefunction (isosurface at 20 % of maximum) of isolated nitrogen (N) and neighboring N–N pair (NN_1) in GaP. Adapted from [472]

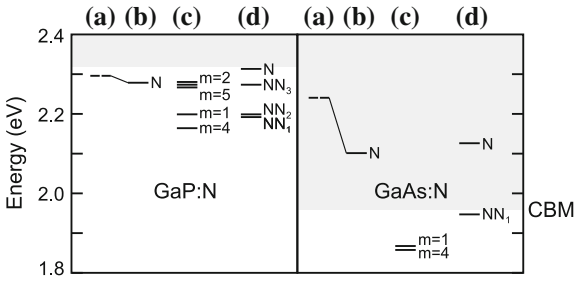

Fig. 7.42 Energy levels of nitrogen impurity states in GaP (*left*) and GaAs (*right*). The energy scale is relative to the bulk GaP valence band maximum, the conduction band minima (CBM) are thus shown relative to the vacuum level. The conduction band is shown in *grey*. For both materials, (a) denotes the isolated N impurity level calculated *without* lattice relaxation (*dashed line*), and (b) *with* lattice relaxation. (c) denotes the position of N–N pair levels, *m* denoting the neighbor. (d) shows selected experimental data. NN_1 denotes the direct neighbor NN-pair. The other NN_n follow the usual nomenclature as in [616]. Data taken from [472] (color figure online)

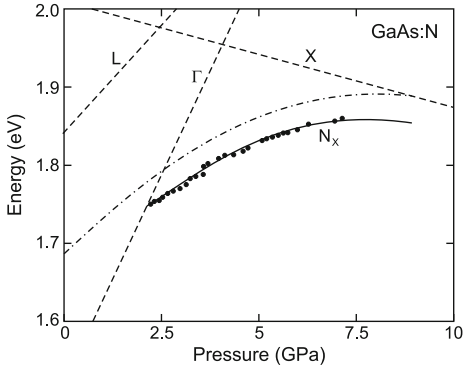

Fig. 7.43 Pressure dependence of the energy of excitons bound to isolated nitrogen impurities in GaAs (*circles*), measured from the top of the GaAs valence band. The *dashed lines* are the pressure dependent GaAs bulk band gaps (cmp. Fig. 6.42). The *solid* (*dash-dotted*) *line* is a theoretical model for the nitrogen-bound exciton (electron) level. Adapted from [615]

(Sects. 9.6.9 and 10.3.2). The wavefunction of an isolated single N impurity and a neighboring N–N pair (NN_1) in GaP are shown in Fig. 7.41b.

Isolated nitrogen impurities in (unstrained) GaAs introduce states only within the conduction band (Fig. 7.42). The reason is that the GaAs conduction band edge is further from the vacuum level than that of GaP (see Fig. 12.21). Only the NN_1 and NN_4 pair levels are theoretically expected to be within the GaAs band gap. The index denotes the *n*-th neighbor position. The NN_1 level has been experimentally observed [613, 614]. The isolated nitrogen impurity level is forced into the GaAs band gap upon hydrostatic pressure [614, 615] (Fig. 7.43). Further levels deeper within the band gap are due to clusters containing more than two nitrogen atoms.

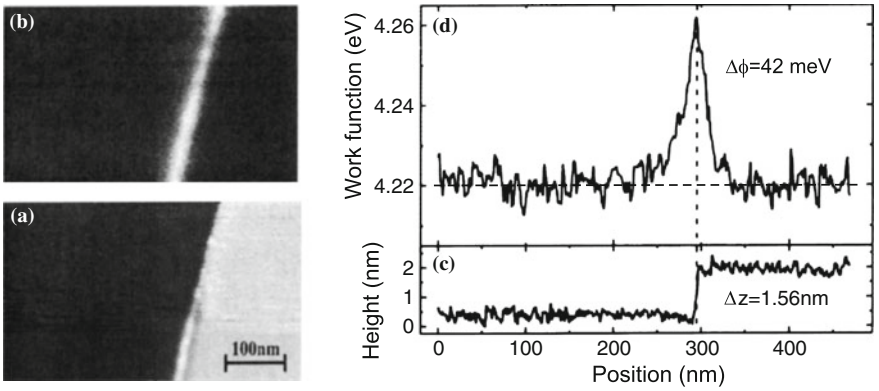

Fig. 7.44 Image of (**a**) topography ($\Delta z = 2.8\,\text{nm}$) and (**b**) work function ($\Delta\phi = 4.21$–$4.26\,\text{eV}$) of a surface step along [111] on a n-GaP(110) surface cleaved in UHV. Adapted from [618]. (**c**) and (**d**) show the corresponding linescans. Adapted from [618]

7.7.10 Surface States

The investigation of (semiconductor) surfaces is a large field with sophisticated methods that allow real-space imaging with atomic resolution by scanning probe microscopy and highly depth resolved electronic studies. The surface represents first of all a break in the periodic crystal potential and thus a defect of the bulk crystal. The unsatisfied bonds partly rearrange, e.g. by building dimers, forming a surface reconstruction or remain as dangling bonds. The surface exhibits a surface density of states. Such states can lie in the band gap and capture electrons, leading to recombination and a depletion layer. For a brief introduction on semiconductor surface physics see Chap. 11; for more details we refer to [617].

As an example of the formation of electronic states at surface defects we show in Fig. 7.44 the comparison of topography and work function (measured by Kelvin probe force microscopy [618]) at a surface step on a GaP(110) surface that has been prepared by cleaving in-situ in ultrahigh vacuum (UHV). The depletion-type band bending of the surface is about 0.4 eV. The further increase of the position of the vacuum level at the step edge shows the presence of trap states in the band gap causing the conduction band to bend upwards (cf. Sect. 21.2.1). Modeling of the effect shows that the charge density at the surface is $6 \times 10^{11}\,\text{cm}^{-2}$ and at the step edge $1.2 \times 10^{6}\,\text{cm}^{-1}$.

7.8 Hydrogen in Semiconductors

The role of hydrogen in semiconductors was first recognized in studies of ZnO [619]. It is now clear that hydrogen plays an important role in the passivation of defects. As a ‘small’ atom, it can attach easily to dangling bonds and form an electron-pair

bond. Thus, surfaces, grain boundaries, dislocations and shallow (donor and acceptor) and deep impurity levels become passivated. A good overview and many details of the physics and technological use of hydrogen in semiconductors can be found in [620, 621]. The hydrogen must be typically introduced as atomic species into semiconductors, e.g. from a plasma in the vicinity of the surface or by ion irradiation.

With regard to silicon it is important to note that the Si–H bond is stronger than the Si–Si bond. Thus a silicon surface under atomic hydrogen exhibits Si–H termination rather than Si–Si dimers [622]. Due to the stronger bond, the hydrogenation leads to an increase of the silicon band gap, which can be used for surface passivation [623], leading to reduced reverse diode current.

The hydrogen concentration in amorphous Si (a-Si) can be as high as 50 % [624]. Electronic grade a-Si contains typically 10–30 atomic % hydrogen and is thus rather a silicon–hydrogen alloy.

Hydrogen in crystalline silicon occupies the bond-center interstitial position (see Fig. 3.18b) as shown in Fig. 7.45a. The complexes formed by hydrogen with shallow acceptors and donors have been studied in detail. It is now generally accepted that for acceptors (e.g. boron) in silicon the hydrogen is located close to the bond-center position of the Si–B pair (BM, bond minimum) as sketched in Fig. 7.46a. The boron atom forms an electron-pair bond with three silicon atoms of the tetrahedra, the fourth silicon bonds to the hydrogen atom. The complex therefore no longer acts as an acceptor. The silicon atoms and the acceptor relax their positions. The adiabatic potential energy surface of hydrogen in Si:B is shown in Fig. 7.45b. The hydrogen can sit on four equivalent sites (BM) along the ⟨111⟩ directions of the initial B–Si_4 tetrahedron. This reduces the symmetry, e.g. of H–B vibrations [626].

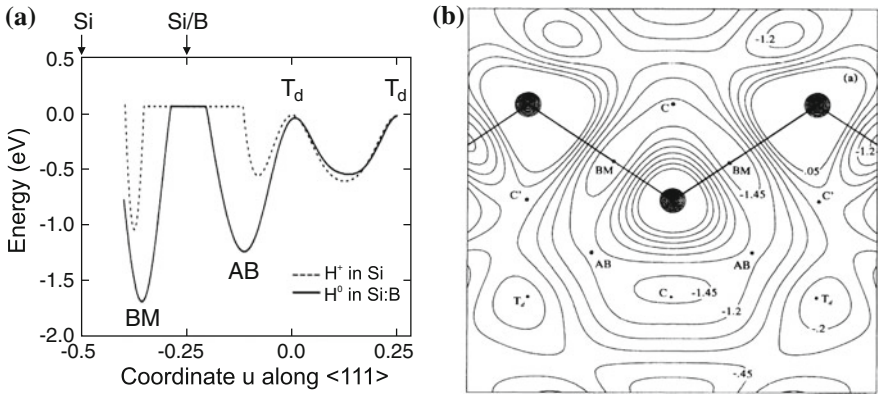

Fig. 7.45 (**a**) Energy for positions u of the hydrogen atom along the ⟨111⟩ direction for H^+ in pure Si (Si atom at $u = -0.25$) and neutral hydrogen (B atom at $u = -0.25$). u is measured in units of $\sqrt{3}a_0$. For all positions of the hydrogen atom the positions of the other atoms have been relaxed in the calculation. Data from [625]. (**b**) Adiabatic potential energy in the (110) plane for hydrogen in Si:B. 'BM' denotes the bond minimum site (high valence electron density), C and C' are equivalent for pure Si. Reprinted with permission from [625], ©1989 APS

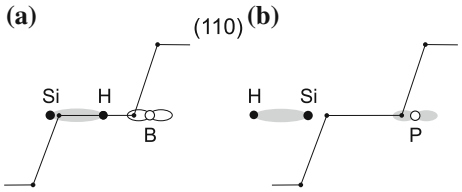

Fig. 7.46 Schematic model for hydrogen in silicon forming a complex with (**a**) a shallow acceptor (boron, empty orbital) and (**b**) a shallow donor (phosphorus, double-filled orbital)

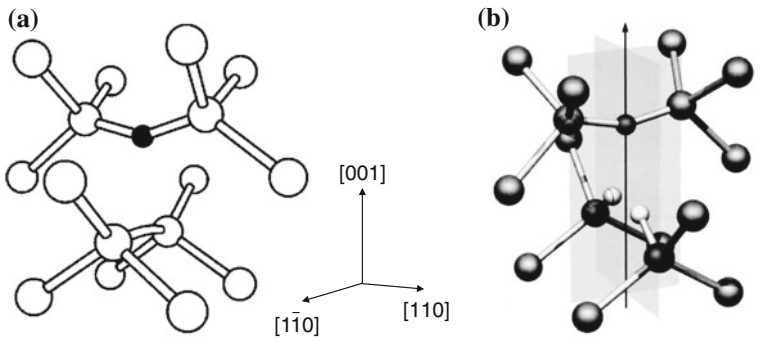

Fig. 7.47 (**a**) Structure of the V–O complex (A center) in silicon. The *black sphere* represents the oxygen atom. Reprinted with permission from [628], ©2004 APS. (**b**) Calculated ground-state structure for the V–O–H_2 center in silicon. Oxygen is over the C_2 axis, and the two *white* spheres represent hydrogen. Reprinted with permission from [629], ©2000 APS

The energetic barrier for the hydrogen orientation has been determined to be 0.2 eV theoretically [625] for a hydrogen motion along the path BM–C–BM in Fig. 7.45b. Stress (along [100] and [112]) reduces the symmetry and leads to splitting of the local vibrational modes, now showing axial symmetry [627]. However, this preferential orientation disappears with an activation energy of 0.19 eV, close to the theoretical value.

Hydrogen has experimentally been found to also passivate shallow donors. The microscopic configuration is sketched in Fig. 7.46b. The hydrogen atom sits on the Si–AB (antibonding) position and forms an electron-pair bond with the silicon atom. The donor, e.g. phosphorus, is left with a double-filled p-orbital (lone pair) whose level is in the valence band and thus no longer contributes to conductivity. Molecular hydrogen can passivate the so-called A center in Si, an oxygen–vacancy complex [629]. The atomistic configuration of the V–O–H_2 complex is shown in Fig. 7.47. The deep double donor S in Si with a level at 0.3 eV below the conduction-band edge can also be passivated by two hydrogen atoms [630].