12 Semiconductors

In Sect. 9.2 we have learned that only a partly filled electronic band can contribute to electric current. Completely filled bands and completely empty bands do not contribute to electrical conductivity and a material which has only completely full and completely empty bands is therefore an insulator. If the distance between the upper edge of the highest filled band (valence band) and the lower edge of the lowest empty band (conduction band) is not too large (e.g. ~ 1 eV), then the finite width of the region over which the Fermi distribution changes rapidly has observable consequences at moderate and high temperatures: A small fraction of the states in the vicinity of the upper edge of the valence band is unoccupied and the corresponding electrons are found in the conduction band. Both these "thermally excited" electrons and the holes that they leave in the valence band can carry electric current. In this case one speaks of a semiconductor. In Fig. 12.1 the differences between a metal, a semiconductor and an insulator are summarized schematically.

A special property of semiconductor materials that is not found in metals is that their electrical conductivity can be altered by many orders of magnitude by adding small quantities of other substances. These additives also determine whether the conductivity is of electron or hole character. The

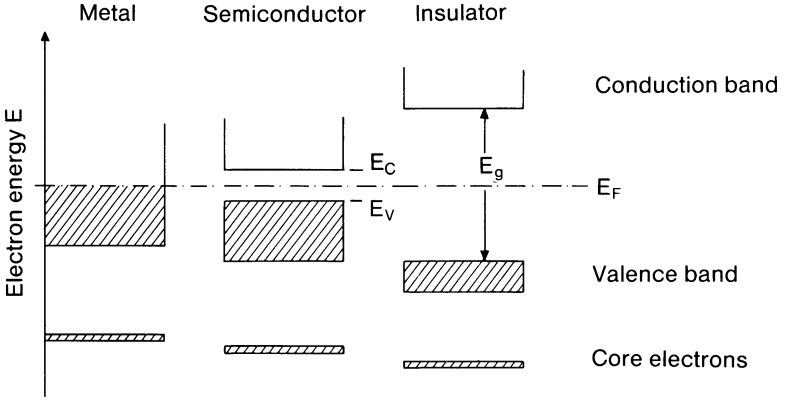

Fig. 12.1. Energy level diagrams for a metal, a semiconductor and an insulator. Metals have a partly occupied band (*shaded*) even at $T = 0$ K. For semiconductors and insulators the Fermi level lies between the occupied valence band and the unoccupied conduction band

H. Ibach, H. Lüth, *Solid-State Physics*, 4th ed.,
DOI 10.1007/978-3-540-93804-0_12,

entire field of solid state electronics relies on this particular property of semiconductors.

12.1 Data of a Number of Important Semiconductors

In Sect. 7.3 we discussed the origin of the band structure of the typical elemental semiconductors diamond (C), Si, and Ge. Due to a mixing of the s- and p-wavefunctions, tetrahedral bonding orbitals (sp^3) are formed, which, for a bonding distance near equilibrium, lead to a splitting into bonding and antibonding orbitals. The bonding orbitals constitute the valence band and the antibonding orbitals the conduction band (Fig. 7.9). Assigning all four s- and p-electrons to the lowest available states leads to a completely filled valence band and a completely empty conduction band. The result is an insulator such as diamond, or, for smaller energy gaps between the valence and conduction band, a semiconductor like silicon or germanium.

From Fig. 7.9 one can deduce an important physical property of the energy gap between the valence and conduction bands: the size of the gap must be temperature dependent. With increasing temperature the lattice parameter increases due to thermal expansion. The splitting between the bonding and antibonding states therefore decreases and the band gap becomes smaller (Table 12.1). A more exact treatment of this effect must also consider the influence of lattice vibrations on E_g. The overall behavior of the band gap with temperature is a linear dependence at room temperature and a quadratic dependence at very low temperature.

Although Fig. 7.9 represents a qualitatively similar picture for the important semiconductors Si and Ge, a quite different picture of the electronic bands emerges in the $E(\boldsymbol{k})$ representation, i.e. in reciprocal space due to the different atomic properties of the bands ($3s$, $3p$ vs. $4s$, $4p$ wavefunctions). These differences can be seen in Fig. 12.2. The curves come from calculations that have been fitted to experimental quantities such as band gap, position of the critical points and effective masses (band curvature).

From this $E(\boldsymbol{k})$ representation along the high symmetry directions in $\boldsymbol{k}$-space, it follows that both semiconductors are so-called indirect-gap

Table 12.1. Energy gaps (width of forbidden bands) between the valence and conduction bands of germanium and silicon

	E_g ($T = 0$ K) [eV]	E_g ($T = 300$ K) [eV]
Si	1.17	1.12
Ge	0.75	0.67

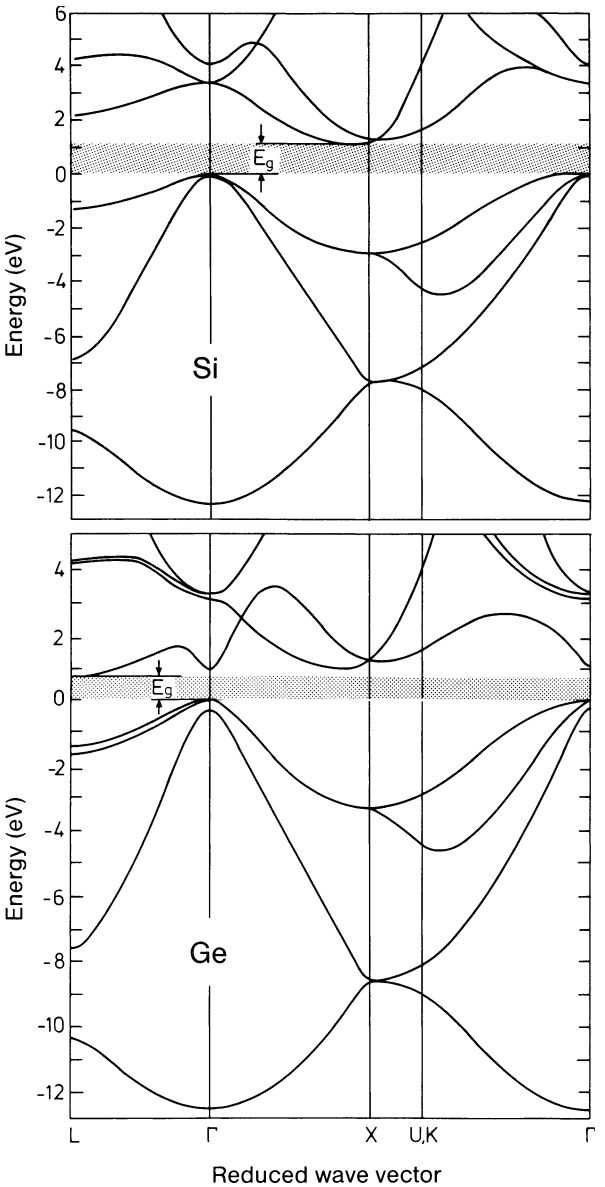

Fig. 12.2. Calculated bandstructures of silicon and germanium. For germanium the spin-orbit splitting is also taken into account. (After [12.1]). Both semiconductors are so-called indirect semiconductors, i.e. the maximum of the valence band and the minimum of the conduction band are at different positions in the Brillouin zone. The minimum of the conduction band of silicon lies along the $\Gamma X = [100]$ direction and that of germanium along the $\Gamma L = [111]$ direction. Note that the form of the Ge bands is very similar to that of Fig. 7.13, although the calculations were performed differently

semiconductors: the minimum distance between the conduction and valence bands (band gap E_g) is between states with different $\boldsymbol{k}$-vectors (Γ, or $\boldsymbol{k} = [000]$ at the valence band maximum of both materials, and $\boldsymbol{k}$ along [111] for Ge, or $\boldsymbol{k}$ along [100] for Si at the conduction band minimum). The conduction band electrons of lowest energy thus have $\boldsymbol{k}$-vectors along the [100] direction in Si, and along the [111] direction in Ge. These regions of $\boldsymbol{k}$-space containing the conduction electrons of Si and Ge are shown in Fig. XV.2.

If $E(\boldsymbol{k})$ is expressed in the parabolic approximation, i.e. retaining terms up to the order k^2, then the surfaces of constant energy are ellipsoids around the [111] or [100] direction. In the principal axis representation (the principal axes are [100] for Si and [111] for Ge), the energy surfaces of the conduction electrons are thus

$$E(\boldsymbol{k}) = \hbar^2 \left(\frac{k_x^2 + k_y^2}{2m_t^*} + \frac{k_z^2}{2m_l^*} \right) = \text{const} . \tag{12.1}$$

Here m_t^* and m_l^* are the "transverse" and "longitudinal" effective masses respectively. The zero of the energy scale is taken at the conduction band minimum.

Measurements by cyclotron resonance of the effective mass of the electrons at these points relative to the mass m of free electrons give the values in Table 12.2.

A detailed study (see Panel XV, "Cyclotron Resonance") of the properties of holes in Si and Ge shows that the structure of the valence band maximum in the vicinity of $\Gamma(\boldsymbol{k} = 0)$ is more complicated than Fig. 12.2 suggests: besides the two valence bands with different curvatures which can be seen in Fig. 12.2, there exists another valence band at Γ which is split off slightly from the other two by $\Delta = 0.29$ eV for Ge and $\Delta = 0.044$ eV for Si. This splitting of the bands stems from spin-orbit interaction, which was not considered in the calculations for Si shown in Fig. 12.2. The qualitative behavior of the bands near Γ is shown in Fig. 12.3 for Si and Ge. In the parabolic approximation one can identify three different effective masses for the holes at Γ which contribute to charge transport. One speaks of heavy and light holes with masses m_{hh}^* or m_{lh}^*, corresponding to the different band curvatures. The holes of the split-off band are called split-off holes and their mass is denoted by m_{soh}^*.

Table 12.2. The transverse (m_t^*) and longitudinal (m_l^*) effective masses relative to the mass m of the free electron for silicon and germanium

	m_t^*/m	m_l^*/m
Si	0.19	0.92
Ge	0.082	1.57

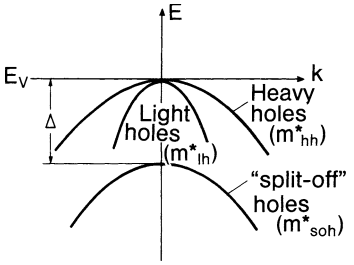

Fig. 12.3. Qualitative bandstructure for Si or Ge in the vicinity of the top of the valence band, including the effect of spin-orbit interaction. Δ is the spin-orbit splitting

Since the formation of sp^3 hybrids is obviously important in the chemical bonding of Si and Ge, one might well expect semiconducting properties in other materials with tetrahedral crystal structure, i.e. with sp^3 hybridization. Based on this consideration, one can correctly identify another important class, the *III–V semiconductors*, which are compound semiconductors comprising elements from the third and fifth groups of the periodic table. Typical examples are InSb, InAs, InP, GaP, GaAs, GaSb, and AlSb. In these compound crystals the bonding is mixed ionic and covalent (cf. Chap. 1). The mixed bonding can be imagined as a superposition of two extreme cases, the *ionic*, in which the electron transfer from Ga to As gives an ionic structure Ga^+As^-, and the *covalent*, in which electron displacement from As to Ga leaves both Ga and As with four electrons in the outer shell, and thereby allows the formation of sp^3 hybrids, just as in Si and Ge. This latter covalent structure is evidently dominant since otherwise the crystal would not be tetrahedrally bonded with the ZnS structure.

In contrast to the elemental semiconductors, the most important representatives of the III–V semiconductors possess a so-called *direct* band gap, i.e., the valence band maximum and conduction band minimum both lie at Γ (Fig. 12.4). There are again three distinct valence bands with a qualitatively similar form at Γ to those of the tetrahedral elemental semiconductors (Fig. 12.3). Important data for a few III–V semiconductors with direct band gaps are summarized in Table 12.3.

For the sake of completeness, we should mention that GaP and AlSb have indirect band gaps similar to Si and Ge (2.32 eV and 1.65 eV, respectively, at $T = 0$ K).

Similar arguments to those presented for the case of III–V compounds lead to an understanding of the so-called II–VI semiconductors, such as ZnO (3.2 eV), ZnS (3.6 eV), CdS (2.42 eV), CdSe (1.74 eV) and CdTe (1.45 eV), where the values in brackets are the direct band gaps E_g at 300 K. In these compounds the bonding is also a mixed ionic-covalent, but now with a larger ionic component than for the III–V semiconductors. The crystal structure is either that of the III–V semiconductors (ZnS) or that of wurtzite (Sect. 2.5). In both cases the local structure is tetrahedral, which can again be attributed to the sp^3 hybridization of the bonding partners.

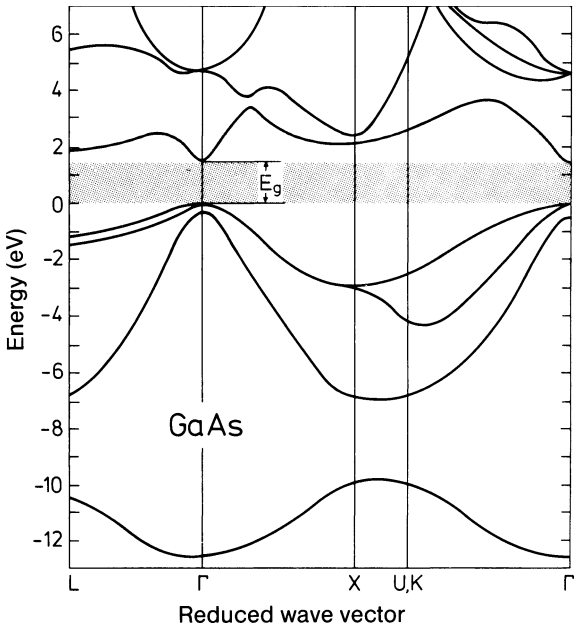

Fig. 12.4. Typical band-structure of a III–V semiconductor, in this case GaAs. (After [12.1])

Table 12.3. Band gap E_g, effective mass m^* and spin-orbit splitting Δ for a few III–V semiconductors: m is the mass of the free electron, m_n^* the effective mass of the electrons, m_{lh}^* the effective mass of the light holes, m_{hh}^* that of the heavy holes, and m_{soh}^* that of the split-off holes

	E_g(0 K) [eV]	E_g(300 K) [eV]	m_n^*/m	m_{lh}^*/m	m_{hh}^*/m	m_{soh}^*/m	Δ [eV]
GaAs	1.52	1.43	0.07	0.08	0.5	0.15	0.34
GaSb	0.81	0.7	0.047	0.05	0.3	0.14	0.8
InSb	0.24	0.18	0.015	0.02	0.4	0.11	0.8
InAs	0.43	0.35	0.026	0.025	0.4	0.14	0.4
InP	1.42	1.35	0.073	0.12	0.6	0.12	0.11

12.2 Charge Carrier Density in Intrinsic Semiconductors

According to the definition of mobility in Sect. 9.5, the electrical conductivity σ of a semiconductor in which electrons and holes contribute to the flow of current, can be written as

$$\sigma = |e|(n\mu_n + p\mu_p) \; . \tag{12.2}$$

Here μ_n and μ_p are the mobilities of the electrons and holes, respectively, and n and p are the corresponding volume concentrations of the charge carriers. The form of (12.2) implies the neglect of any energy dependence

($\boldsymbol{k}$-dependence) of the quantities μ_n and μ_p in a first approximation. This is because it is generally sufficient to consider only charge carriers in the parabolic part of the bands, where the effective mass approximation is valid (i.e. m_n^* and m_p^* are constant). Because of the opposite signs of both the drift velocity and the electrical charge e for holes and electrons, both types of charge carrier contribute with the *same* sign to σ.

In contrast to metallic conductivity, the conductivity of semiconductors is strongly temperature dependent. This is because the band gap E_g, across which "free" charge carriers must be thermally excited, causes (via the Fermi distribution – see below) a strong dependence of the charge carrier concentrations n and p on the temperature.

Semiconductors are called *intrinsic* when "free" electrons and holes can be created only by electronic excitations from the valence band to the conduction band. (In Sect. 12.3 we will also consider excitations from defects and impurities.) As in any solid, the occupation of the energy levels in semiconductors must obey Fermi statistics $f(E, T)$ (Sect. 6.3), i.e.

$$n = \int_{E_c}^{\infty} D_C(E) f(E, T) dE \,, \tag{12.3 a}$$

and for holes

$$p = \int_{-\infty}^{E_v} D_V(E)[1 - f(E, T)] dE \,. \tag{12.3 b}$$

The ranges of integration should really extend only to the upper and lower band edges, respectively. However, because the Fermi function $f(E, T)$ decreases sufficiently rapidly, the ranges can be extended to infinity. The functions $D_C(E)$ and $D_V(E)$ are the densities of states in the conduction and valence bands, respectively. In the parabolic approximation ($m^* = \text{const}$) one has [cf. (6.11)]:

$$D_C(E) = \frac{(2m_n^*)^{3/2}}{2\pi^2\hbar^3} \sqrt{E - E_C} \,, \quad (E > E_C) \,; \tag{12.4 a}$$

$$D_V(E) = \frac{(2m_p^*)^{3/2}}{2\pi^2\hbar^3} \sqrt{E_V - E} \,, \quad (E < E_V) \,. \tag{12.4 b}$$

The density in the region of the forbidden band $E_V < E < E_C$ is of course zero. In an intrinsic semiconductor all "free" electrons in the conduction band originate from states in the valence band, and thus the concentration of holes p must be equal to the concentration of "free" electrons n. The situation is sketched in Fig. 12.5. If the effective masses m_n^* and m_p^* and therefore also the densities of states D_C and D_V are equal, the Fermi level E_F must lie in the middle of the forbidden band. If D_C and D_V are different

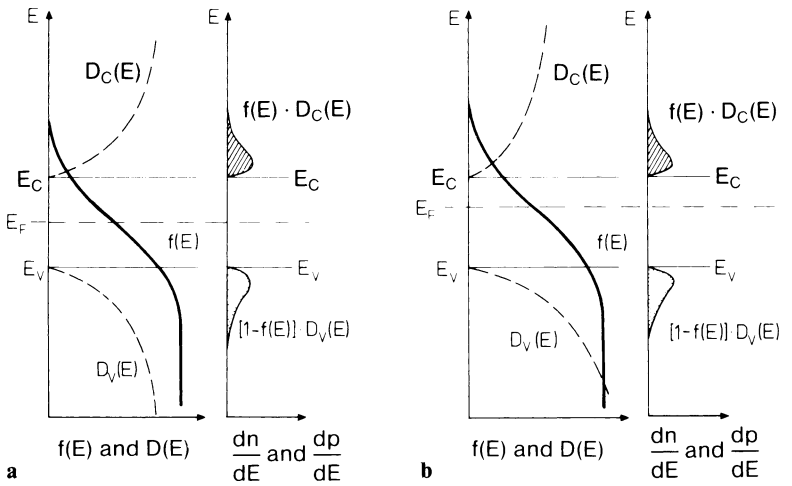

Fig. 12.5. (**a**) Fermi function $f(E)$, density of states $D(E)$ and electron (n) and hole (p) concentrations in the conduction and valence bands for the case of equal densities of states in the conduction and valence bands (schematic); (**b**) the same figure for the case of differing densities of states in the conduction and valence bands. The number of holes must again be equal to the number of electrons, and thus the Fermi level no longer lies in the middle of the gap between conduction and valence bands; its position then becomes temperature dependent

from one another, then E_F shifts slightly towards one of the band edges, such that the occupation integrals (12.3) remain equal.

Because the "width" of the Fermi function ($\sim 2\not{k}T$) is, at normal temperatures, small compared to the gap width (~ 1 eV), the Fermi function $f(E,T)$ can be approximated by Boltzmann occupation statistics within the bands ($E > E_C$ and $E < E_V$), i.e. for the conduction band:

$$\frac{1}{\exp[(E-E_F)/\not{k}T]+1} \sim \exp\left(-\frac{E-E_F}{\not{k}T}\right) \ll 1 \quad \text{for} \quad E-E_F \gg 2\not{k}T\,. \tag{12.5}$$

From (12.3a) and (12.4a), the electron concentration n in the conduction band follows as

$$n = \frac{(2m_n^*)^{3/2}}{2\pi^2\hbar^3} e^{E_F/\not{k}T} \int_{E_C}^{\infty} \sqrt{E-E_C} \cdot e^{-E/\not{k}T} dE\,. \tag{12.6}$$

Substituting $X_C = (E–E_C)/\not{k}T$, we obtain the formula

$$n = \frac{(2m_n^*)^{3/2}}{2\pi^2\hbar^3} (\not{k}T)^{3/2} \exp\left(-\frac{E_C-E_F}{\not{k}T}\right) \int_0^{\infty} X_C^{1/2} e^{-X_C} dX_C\,. \tag{12.7}$$

With an analogous calculation for the valence band, one finally obtains the expressions

$$n = 2\left(\frac{2\pi m_n^* \not{k} T}{h^2}\right)^{3/2} \exp\left(-\frac{E_C - E_F}{\not{k} T}\right) = N_{eff}^C \exp\left(-\frac{E_C - E_F}{\not{k} T}\right), \tag{12.8a}$$

$$p = 2\left(\frac{2\pi m_p^* \not{k} T}{h^2}\right)^{3/2} \exp\left(\frac{E_V - E_F}{\not{k} T}\right) = N_{eff}^V \exp\left(\frac{E_V - E_F}{\not{k} T}\right). \tag{12.8b}$$

The prefactors N_{eff}^C and N_{eff}^V are the well-known partition functions for translational motion in three dimensions. One sees that small concentrations of free charge carriers in semiconductors can be approximately described by Boltzmann statistics (the approximation to Fermi statistics for E–$E_F \gg 2 \not{k} T$). If one compares Fig. 12.5 with the potential well model of Chap. 6, the conduction band can be formally regarded as a potential well, in which the Fermi level E_F lies well below ($\gg \not{k} T$) the bottom of the potential well.

The use of the so-called "effective densities of states" N_{eff}^C and N_{eff}^V in (12.8) allows a further formal interpretation in which the whole conduction (valence) band can be characterized by a single energy level E_C (E_V) (i.e., the band edge) with the density of states N_{eff}^C (N_{eff}^V) (temperature dependent!). The occupation densities n and p of these bands are determined by the Boltzmann factor (with the energy in each case measured from E_F). This approximation, which is often valid for semiconductors, is called the *approximation* of non-degeneracy. High densities of charge carriers can be created in semiconductors by means of high densities of impurities (Sect. 12.3). The approximation of non-degeneracy is then no longer valid and one speaks of *degenerate semiconductors*.

From (12.8a, b) the following generally valid relationship can be derived ($E_g = E_C$–E_V):

$$n\,p = N_{eff}^C N_{eff}^V e^{-E_g/\not{k} T} = 4\left(\frac{\not{k} T}{2\pi\hbar^2}\right)^3 (m_n^* m_p^*)^{3/2} e^{-E_g/\not{k} T}. \tag{12.9}$$

This equation implies that for a particular semiconductor, which is completely characterized by its absolute band gap E_g and the effective masses m_n^* and m_p^* in the conduction and valence bands, the electron and hole concentrations behave as a function of temperature according to the *law of mass action*.

If we further assume an intrinsic semiconductor ($n = p$), then the intrinsic charge carrier concentration n_i varies with temperature as follows

$$n_i = p_i = \sqrt{N_{eff}^C N_{eff}^V}\, e^{-E_g/2\not{k} T} = 2\left(\frac{\not{k} T}{2\pi\hbar^2}\right)^{3/2} (m_n^* m_p^*)^{3/4} e^{-E_g/2\not{k} T}. \tag{12.10}$$

Values of n_i and E_g for the important materials Ge, Si and GaAs are summarized in Table 12.4.

Table 12.4. Band gap E_g and the intrinsic carrier concentration n_i for germanium, silicon and gallium arsenide at 300 K

	E_g [eV]	n_i [cm^{-3}]
Ge	0.67	2.4×10^{13}
Si	1.1	1.5×10^{10}
GaAs	1.43	5×10^{7}

According to (12.8a, b), the Fermi level at a particular temperature adopts the position necessary to yield charge neutrality, i.e.

$$n = p = N_{\text{eff}}^{\text{C}} e^{-E_C/\not{k}T} e^{E_F/\not{k}T} = N_{\text{eff}}^{\text{V}} e^{E_V/\not{k}T} e^{-E_F/\not{k}T}, \tag{12.11}$$

$$e^{2E_F/\not{k}T} = \frac{N_{\text{eff}}^{\text{V}}}{N_{\text{eff}}^{\text{C}}} e^{(E_V+E_C)/\not{k}T}, \tag{12.12}$$

$$E_F = \frac{E_C + E_V}{2} + \frac{\not{k}T}{2} \ln(N_{\text{eff}}^{\text{V}}/N_{\text{eff}}^{\text{C}}) = \frac{E_C + E_V}{2} + \frac{3}{4}\not{k}T \ln(m_p^*/m_n^*) . \tag{12.13}$$

If the effective densities of states and effective masses (i.e. the band curvature of the conduction and valence bands) are equal, then the Fermi level of an intrinsic semiconductor lies exactly in the middle of the forbidden band, and this is true for all temperatures. If, however, the effective densities of states in the conduction and valence bands are different, then the Fermi function lies asymmetrically with respect to the band edges E_C and E_V (Fig. 12.5) and the Fermi level shows a weak temperature dependence according to (12.13).

12.3 Doping of Semiconductors

The intrinsic carrier concentration n_i of 1.5×10^{10} cm^{-3} (at 300 K) of Si is not nearly large enough to yield the current densities necessary for practical semiconductor devices. Concentrations that are orders of magnitude higher than n_i can be created by doping, i.e. by the addition of electrically active impurities to the semiconductor. Most semiconductors cannot be grown as single crystals with sufficient purity that one can observe intrinsic conductivity at room temperature. Unintentional doping, even in the purest commercially available GaAs single crystals, leads to carrier densities of about 10^{16} cm^{-3} (at 300 K), compared with a corresponding intrinsic concentration of $n_i = 5 \times 10^7$ cm^{-3}.

Electrically active impurities in a semiconductor raise the concentration of either the "free" electrons or the "free" holes by donating electrons to the conduction band or by accepting them from the valence band. These

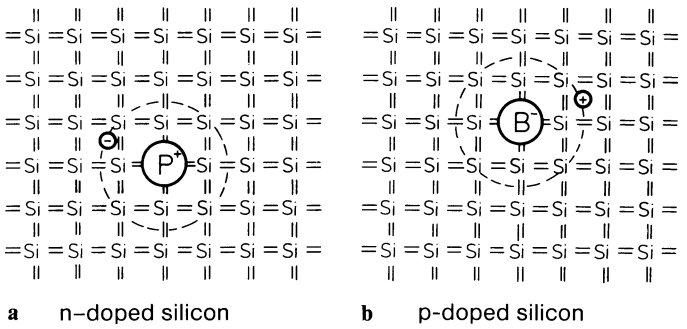

Fig. 12.6 a, b. Schematic representation of the effect of a donor (**a**) and an acceptor (**b**) in a silicon lattice. The valence-five phosphorus atom is incorporated in the lattice at the site of a silicon atom. The fifth valence electron of the phosphorus atom is not required for bonding and is thus only weakly bound. The binding energy can be estimated by treating the system as a hydrogen atom embedded in a dielectric medium. The case of an acceptor (**b**) can be described similarly: the valence-three boron accepts an electron from the silicon lattice. The hole that is thereby created in the valence band orbits around the negatively charged impurity. The lattice constant and the radius of the defect center are not drawn to scale. In reality the first Bohr radius of the "impurity orbit" is about ten times as large as the lattice parameter

impurities are called *donors* and *acceptors*, respectively. A donor in a Si lattice is created, for example, when a valence-four Si atom is replaced by a valence-five atom such as P, As or Sb. The electronic structure of the outermost shell of these impurities is s^2p^3 instead of the $3s^2 3p^2$ Si configuration. To adopt the tetrahedral bonding structure of an sp^3 hybrid in the lattice, only the s^2p^2 electrons of the valence-five atom are necessary; the excess electron in the p shell has no place in the sp^3 hybrid bond. This electron can be imagined as weakly bound to the positively charged and tetrahedrally bonded donor core, which has replaced a Si atom in the lattice (Fig. 12.6 a).

A valence-five donor impurity in the Si lattice can be pictured, to a good approximation, as a positively charged monovalent core, to which one electron is bound. The electron can become dissociated and can then move "freely" through the lattice; i.e., on ionization, this electron will be excited from the impurity into the conduction band. The donor impurity can be described as a hydrogen-like center, in which the Coulomb attraction between the core and the valence electron is screened by the presence of Si electrons in the vicinity.

To estimate the excitation and ionization energy of the extra phosphorus electron (Fig. 12.6 a), one can approximate the screening effect of the surrounding Si by inserting the dielectric constant of Si ($\varepsilon_{\mathrm{Si}} = 11.7$) into the expression for the energy levels of the hydrogen atom. The energy levels for the Rydberg series of the hydrogen atom are

$$E_n^{\mathrm{H}} = \frac{m_{\mathrm{e}}\, e^4}{2(4\pi\varepsilon_0\hbar)^2}\,\frac{1}{n^2} \tag{12.14}$$

and for the $n = 1$ level, the ionization energy is 13.6 eV. For the P donor, the mass m_e of the free electron must be replaced by the effective mass $m_n^* = 0.3\, m_e$ of a silicon conduction electron, and the dielectric constant ε_0 of vacuum by $\varepsilon_0 \varepsilon_{Si}$. A value for the ionization energy E_d of the donor of ~ 30 meV results. The energy level E_D of the donor electron in the bound state should thus lie about 30 meV below the conduction band edge E_C. A still smaller value is obtained for germanium. Here, $\varepsilon_{Ge} = 15.8$ and $m_n^* \sim 0.12\, m_e$. An estimate of $(E_C - E_D) \simeq 6$ meV is obtained. The situation is depicted in a bandstructure scheme in Fig. 12.7, where only the ground state of the donor is drawn. Between this ground state and the conduction band edge are a series of excited states [$n > 1$ in (12.14)], whose spacing decreases with increasing energy, and which finally join the continuum of the conduction band. The situation is very similar to that of the H atom, where the conduction band continuum corresponds to the unbound states above the vacuum level. The energy of the excited states can be determined, for example, from optical spectra. Figure 12.8 shows an absorption spectrum of the Sb donor in Ge. The bands below 9.6 meV correspond to excitations from the ground state to higher, excited states. The spectrum is actually more complicated than would be expected from the simple hydrogenic model; this is because the crystal field lifts the partial degeneracy of the hydrogen-like states. Above a photon energy of 9.6 meV, the electrons are excited to the continuum of the conduction band.

As can be seen from the experimental example in Fig. 12.8, the simple description of the donor by a hydrogenic model allows one to estimate the order of magnitude of the ionization energy E_d. Within this model all donor impurities, such as P, As and Sb, should have the same ionization energy E_d when present in the same semiconductor host. The experimental values of E_d in Table 12.5 show, however, that the values vary somewhat from donor to donor.

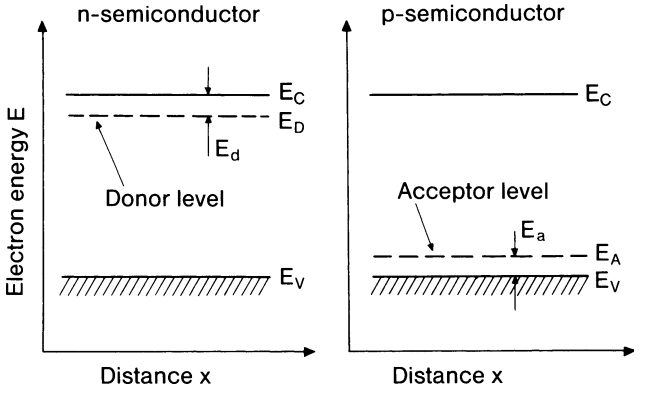

Fig. 12.7. Qualitative position of the ground state levels of donors and acceptors relative to the minimum of the conduction band E_C and the maximum of the valence band E_V. The quantities E_d and E_a are the ionization energies of the donor and acceptor, respectively

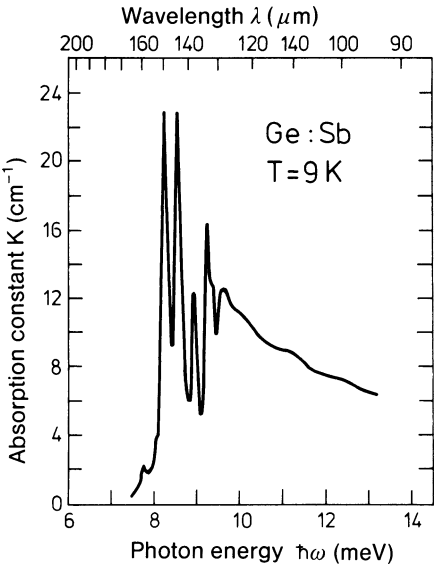

Fig. 12.8. Optical absorption spectrum of a Sb donor in germanium, measured at $T = 9$ K. (After [12.2])

Table 12.5. Ionization energies E_d for a few donor species in silicon and germanium

	P [meV]	As [meV]	Sb [meV]
Si	45	54	43
Ge	13	14	10

It is not surprising to find that the crude description of screening in terms of the dielectric constant of the semiconductor is inadequate to describe the finer details resulting from atomic effects.

That the description by a macroscopic dielectric constant nonetheless yields good values for E_d, is because the screening leads to a wavefunction which is "smeared out" over many lattice constants. Inserting the dielectric constant of the semiconductor ε_s in the formula for the Bohr radius

$$r = \varepsilon_0 \varepsilon_s \frac{h^2}{\pi m_n^* e^2} \tag{12.15}$$

expands this radius by a factor of ε_s (~ 12 for Si) compared to the (hydrogen) Bohr-radius.

The bound valence electron of the donor impurity is therefore "smeared" over about 10^3 lattice sites.

If a valence-three impurity atom (B, Al, Ga, In) is present in the lattice of a valence-four elemental semiconductor (Si, Ge), the sp^3 hybrid responsible for the tetrahedral bonding can easily accept an electron from the valence band

Table 12.6. Ionization energies E_a for a few acceptor species in silicon and germanium

	B [meV]	Al [meV]	Ga [meV]	In [meV]
Si	45	67	74	153
Ge	11	11	11	12

and leave behind a hole (Fig. 12.6 b). Such impurities are called *acceptors*. An acceptor is neutral at very low temperature. It becomes ionized when an electron obtains sufficient energy to be lifted from the valence band to the so-called acceptor level. Acceptors therefore have the charge character "from neutral to negative", while donors are ionized from neutral to positive. The hole that exists in the vicinity of an ionized acceptor is in the screened Coulomb field of the fixed, negative impurity, and the energy needed to separate the hole from the ion and create a "free" hole in the valence band can be estimated using the hydrogenic model as in the case of donors. The relationships derived for donors and acceptors are fundamentally similar except for the sign of the charge. Table 12.6 shows that the ionization energies E_a for acceptors are in fact quite close to those of donors. Semiconductors doped with donors and acceptors are called *n*- and *p*-type materials, respectively.

The lowest impurity concentrations that can presently be achieved in semiconductor single crystals are of the order of $10^{12}\,\mathrm{cm}^{-3}$. Ge with an intrinsic charge carrier concentration n_i of $2.4 \times 10^{13}\,\mathrm{cm}^{-3}$ (at 300 K) is therefore obtainable as intrinsic material at room temperature, whereas Si ($n_i = 1.5 \times 10^{10}\,\mathrm{cm}^{-3}$ at 300 K) does not show intrinsic conductivity at room temperature. In addition to the electrically active impurities discussed here, a semiconductor may of course contain many impurities and defects that cannot be ionized as easily and thus do not affect the electrical conductivity.

12.4 Carrier Densities in Doped Semiconductors

In a doped semiconductor, an electron in the conduction band can originate either from the valence band or from the ionization of a donor; likewise, a hole in the valence band may correspond either to the electron in the conduction band or to a negatively charged (ionized) acceptor. For a non-degenerate semiconductor, the occupation of the conduction and valence bands must nonetheless be governed by the Boltzmann approximation (12.8 a, b). Therefore the so-called "law of mass action" must also be valid for the doped semiconductor (12.9)

$$n\,p = N_{\mathrm{eff}}^{\mathrm{C}} N_{\mathrm{eff}}^{\mathrm{V}} \mathrm{e}^{-E_g/\not{k}\,T}\,,$$

in which the position of the Fermi level E_F no longer appears. In comparison with an intrinsic semiconductor, the position of E_F is now governed by a rather more complicated "neutrality condition", which also takes account of the charge on the impurities: the terms used in the following discussion are represented schematically in Fig. 12.9. The density of all available donors N_D and acceptors N_A is composed of the density of the neutral donors and acceptors, N_D^0 and N_A^0, and the density of ionized donors N_D^+ (which are then positively charged) and ionized acceptors N_A^- (negatively charged). In a homogeneous semiconductor, the negative charge density $n + N_A^-$ must be compensated by an equally large positive charge density $p + N_D^+$ (see Fig. 12.9). Hence, the following neutrality condition governs the position of the Fermi level E_F in a homogeneously doped semiconductor:

$$n + N_A^- = p + N_D^+ , \tag{12.16}$$

in which

$$N_D = N_D^0 + N_D^+ , \tag{12.17 a}$$

$$N_A = N_A^0 + N_A^- . \tag{12.17 b}$$

For typical impurity concentrations (10^{13}–$10^{17}\,cm^{-3}$), in which the individual donors and acceptors do not influence one another, the occupation of the donors by electrons (n_D) and of acceptors by holes (p_A) is given to a good approximation by

$$n_D = N_D^0 = N_D[1 + \exp(E_D - E_F)/kT]^{-1} , \tag{12.18 a}$$

$$p_A = N_A^0 = N_A[1 + \exp(E_F - E_A)/kT]^{-1} . \tag{12.18 b}$$

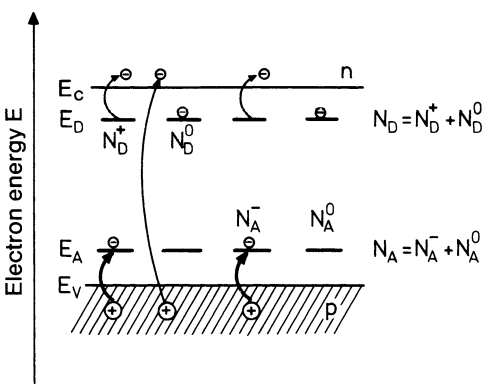

Fig. 12.9. Explanation of the notation commonly used for carrier and impurity concentrations in *n*- and *p*-type semiconductors: n and p are the concentrations of "free" electrons and holes. The total concentrations N_D and N_A of donors and acceptors consist of the density of neutral, N_D^0 or N_A^0, and ionized, N_D^+ or N_A^-, donors and acceptors, respectively. Electrons in the conduction band (density n) and holes in the valence band (density p) originate either from interband excitations or from impurities

We neglect here a modification of the Fermi function (multiplication of the exponential term by $\frac{1}{2}$), which takes into account the possibility of the capture of a single electron only, however, with two distinguishable spin states.

The general case, in which both donors and acceptors are considered simultaneously, can only be treated numerically. We therefore restrict the present treatment to the case of a pure *n-type semiconductor*, in which only donors are available. Equations (12.8 a, 12.17 a, 12.18 a) can then be used to calculate the concentration of charge carriers. For convenience these are repeated below:

$$n = N_{\mathrm{eff}}^{\mathrm{C}} \mathrm{e}^{-(E_{\mathrm{C}}-E_{\mathrm{F}})/\not{k}\,T} , \tag{12.8 a}$$

$$N_{\mathrm{D}} = N_{\mathrm{D}}^{0} + N_{\mathrm{D}}^{+} , \tag{12.17 a}$$

$$N_{\mathrm{D}}^{0} = N_{\mathrm{D}}[1 + \exp(E_{\mathrm{D}} - E_{\mathrm{F}})/\not{k}\,T]^{-1} . \tag{12.18 a}$$

"Free" electrons in the conduction band can only originate from donors or from the valence band, i.e.,

$$n = N_{\mathrm{D}}^{+} + p . \tag{12.19}$$

As a further simplification we assume that the main contribution to the conductivity stems from ionized donors, i.e. that $N_{\mathrm{D}}^{+} \gg n_{\mathrm{i}}$ ($np = n_{\mathrm{i}}^2$). For Si for example ($n_{\mathrm{i}} = 1.5 \times 10^{10}\,\mathrm{cm}^{-3}$ at 300 K) this is readily fulfilled even at low doping levels. For this simple case, (12.19) is replaced by

$$n \approx N_{\mathrm{D}}^{+} = N_{\mathrm{D}} - N_{\mathrm{D}}^{0} , \tag{12.20}$$

i.e. with (12.18 a) one has

$$n \approx N_{\mathrm{D}} \left(1 - \frac{1}{1 + \exp[(E_{\mathrm{D}} - E_{\mathrm{F}})/\not{k}\,T]} \right) . \tag{12.21}$$

With the help of (12.8 a), E_{F} can be expressed via

$$(n/N_{\mathrm{eff}}^{\mathrm{C}}) \mathrm{e}^{E_{\mathrm{C}}/\not{k}\,T} = \mathrm{e}^{E_{\mathrm{F}}/\not{k}\,T} , \tag{12.22}$$

to yield

$$n \approx \frac{N_{\mathrm{D}}}{1 + \mathrm{e}^{E_{\mathrm{d}}/\not{k}\,T} n/N_{\mathrm{eff}}^{\mathrm{C}}} , \tag{12.23}$$

where $E_{\mathrm{d}} = E_{\mathrm{C}} - E_{\mathrm{D}}$ is the energetic distance of the donor level from the conduction band edge. Expressing (12.23) as a quadratic equation for n we have

$$n + \frac{n^2}{N_{\mathrm{eff}}^{\mathrm{C}}} \mathrm{e}^{E_{\mathrm{d}}/\not{k}\,T} \approx N_{\mathrm{D}} . \tag{12.24}$$

The physically meaningful solution is

$$n \approx 2N_{\mathrm{D}} \left(1 + \sqrt{1 + 4 \frac{N_{\mathrm{D}}}{N_{\mathrm{eff}}^{\mathrm{C}}} \mathrm{e}^{E_{\mathrm{d}}/\not{k}T}} \right)^{-1} . \tag{12.25}$$

This expression for the conduction electron concentration in n-type semiconductors contains the following limiting cases:

I) If the temperature T is so low that

$$4(N_{\mathrm{D}}/N_{\mathrm{eff}}^{\mathrm{L}})\mathrm{e}^{E_{\mathrm{d}}/\not{k}T} \gg 1 \tag{12.26}$$

then one has

$$n \approx \sqrt{N_{\mathrm{D}} N_{\mathrm{eff}}^{\mathrm{C}}} \mathrm{e}^{-E_{\mathrm{d}}/2\not{k}T} . \tag{12.27}$$

In this region, where a sufficiently large number of donors still retain their valence electrons, i.e. are not ionized, one speaks of the *freeze-out range* of carriers. The similarity between (12.27) and (12.10) should be noted: instead of the valence band quantities $N_{\mathrm{eff}}^{\mathrm{V}}$ and E_{V}, one now has the corresponding donor quantities N_{D} and E_{d} (i.e. E_{d} instead of E_{g}). In this low-temperature regime, the electron concentration depends exponentially on the temperature T as in an intrinsic semiconductor; here, however, it is the much smaller donor ionization energy E_{d} that appears in place of E_{g}. The special case treated here of a pure n-type doping is of course rarely realized in practice. Trace quantities of acceptors are almost always present. As a consequence, the Fermi level lies below E_{D}, and an activation energy of E_{d} is therefore found in most cases.

II) At temperatures T, for which

$$4(N_{\mathrm{D}}/N_{\mathrm{eff}}^{\mathrm{C}})\mathrm{e}^{E_{\mathrm{d}}/\not{k}T} \ll 1 , \tag{12.28}$$

Eq. (12.25) becomes

$$n \approx N_{\mathrm{D}} = \mathrm{const} , \tag{12.29}$$

i.e. the concentration of donor electrons in the conduction band has reached the maximum possible value, equal to the concentration of donors; all donors are ionized, and one speaks of the *saturation range*. In a first approximation, one may still neglect electrons excited from the valence band.

III) At yet higher temperatures the concentration of electrons excited from the valence band across the gap E_{g} increases, and eventually outweighs the electron density due to donors. In this region the n-type material behaves as an intrinsic semiconductor and one speaks of the *intrinsic region* of the carrier concentration. The various temperature and carrier concentration regimes are portrayed in Fig. 12.10 together with the corresponding position of the Fermi level E_{F}.

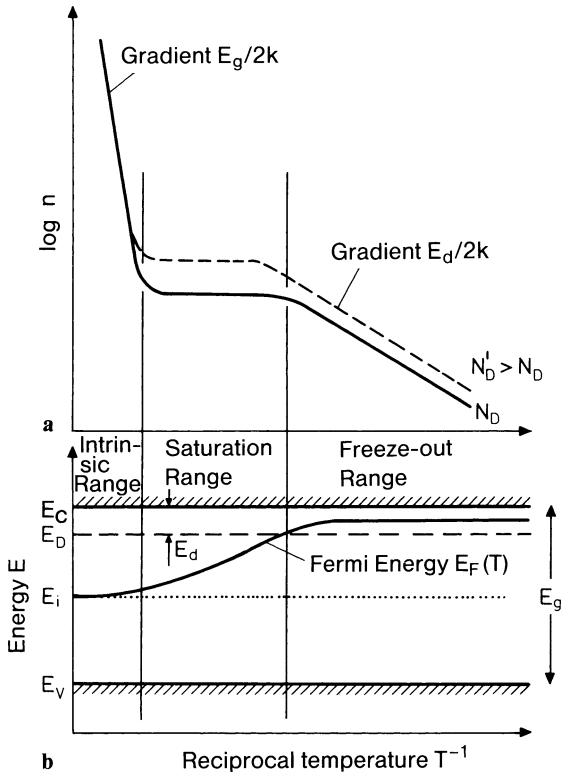

Fig. 12.10. (**a**) Qualitative temperature dependence of the concentration n of electrons in the conduction band of an n-type semiconductor for two different donor concentrations $N_D' > N_D$. The width of the forbidden band is E_g and E_d is the ionization energy of the donors; (**b**) qualitative temperature dependence of the Fermi energy $E_F(T)$ in the same semiconductor. E_C and E_V are the lower edge of the conduction band and the upper edge of the valence band, respectively, E_D is the position of the donor levels and E_i is the Fermi level of an intrinsic semiconductor

The position of the Fermi level E_F as a function of temperature can be discussed analogously to the case of an intrinsic semiconductor, but it will not be explicitly worked out here. In the freeze-out range, one may replace the valence band edge by the donor level. In the region of very high temperature, i.e. in the intrinsic region, the laws pertinent to an intrinsic semiconductor apply.

For n-doped Si with a phosphorus concentration of $3 \times 10^{14}\,\mathrm{cm}^{-3}$, the saturation range stretches between 45 and 500 K, i.e., at room temperature all donors are ionized. Figure 12.11 shows experimental results for the electron concentration $n(T)$, determined by Hall effect measurements, of n-doped Ge with impurity concentrations between 10^{13} and $10^{18}\,\mathrm{cm}^{-3}$. The relationships sketched qualitatively in Fig. 12.10 are clearly recognizable.

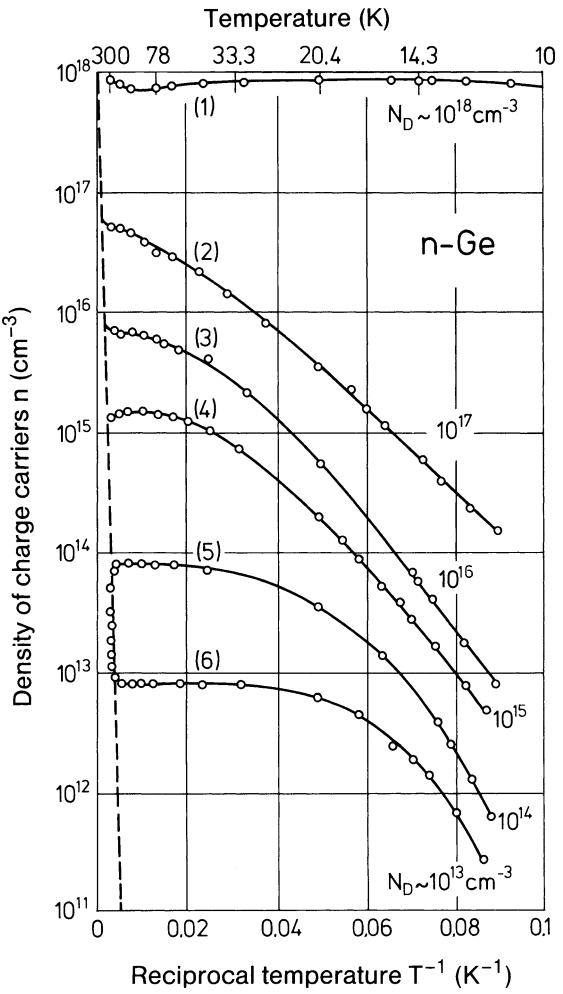

Fig. 12.11. The concentration n of free electrons in n-type germanium, measured using the Hall effect (Panel XIV). For the samples (*1*) to (*6*), the donor concentration N_D varies between 10^{18} and $10^{13}\,cm^{-3}$. The temperature dependence of the electron concentration in the intrinsic region is shown by the dashed line. (After [12.3])

12.5 Conductivity of Semiconductors

As was discussed more fully in Sect. 9.5, a calculation of the electrical conductivity σ as a function of temperature T requires a more detailed treatment of the mobility. For a semiconductor, one must consider both the electrons in the lower conduction band region (concentration: n, mobility: μ_n) and the holes at the upper valence band edge (p, μ_p). The current density $\boldsymbol{j}$ in an isotropic semiconductor (non-tensorial conductivity σ) is therefore

$$\boldsymbol{j} = e(n\mu_n + p\mu_p)\boldsymbol{\mathscr{E}} \,. \tag{12.30}$$

In contrast to a metal, where only electrons at the Fermi edge need be considered, i.e. $\mu = \mu(E_F)$ (Sect. 9.5), the mobilities μ_n and μ_p in a semiconductor are average values for the electron and hole occupied states in the lower conduction band and upper valence band, respectively. In non-degenerate semiconductors, Fermi statistics can be approximated by Boltzmann statistics in this region. Within the framework of this approximation, the averaging, which will not be carried out in more detail, yields the following expression

$$\mu_n = \frac{1}{m_n^*} \frac{\langle \tau(\boldsymbol{k}) v^2(\boldsymbol{k}) \rangle}{\langle v^2(\boldsymbol{k}) \rangle} e \,. \tag{12.31}$$

Here $v(\boldsymbol{k})$ is the velocity of an electron in an electric field $\mathscr{E}$ at the point $\boldsymbol{k}$ in reciprocal space, and $\tau(\boldsymbol{k})$ is its relaxation time (a more exact definition is given in Sect. 9.4). For holes, the quantities in (12.31) are taken at corresponding points in the valence band, and the effective mass m_n^* of the electrons is replaced by that of holes.

Instead of deriving a rigorous solution of the Boltzmann equation and an exact further treatment of (12.31), we restrict ourselves in the present case, as we did for metals (Sect. 9.5), to a more qualitative discussion of the scattering processes in semiconductors. In this respect, electrons and holes behave in a qualitatively similar way. After considerable simplification, (12.31) yields a proportionality between the mobility and the relaxation time ($\mu \propto \tau$). For metals, this proportionality holds exactly, see (9.58 b). Because τ is also proportional to the average time between collisions, it follows that

$$\frac{1}{\tau} \propto \langle v \rangle \Sigma \,, \tag{12.32}$$

where Σ represents the scattering cross section for electrons and holes at a scattering center. In contrast to the case of metals (Sect. 9.5), $\langle v \rangle$ is to be considered as a thermal average (according to Boltzmann statistics) over all electron or hole velocities in the lower conduction band or upper valence band, respectively. Because of the validity of Boltzmann statistics in semiconductors, we have

$$\langle v \rangle \propto \sqrt{T} \,. \tag{12.33}$$

If we now consider *scattering from acoustic phonons*, we can estimate the cross section Σ_{ph}, as in the case of metals (Sect. 9.5), from the square of the average vibrational amplitude $\langle u^2(\boldsymbol{q}) \rangle$ of a phonon $(\boldsymbol{q}, \omega_{\boldsymbol{q}})$, i.e. for temperatures $T \gg \Theta$ (Θ is the Debye temperature), we have (Sect. 9.5)

$$M \omega_{\boldsymbol{q}}^2 \langle u^2(\boldsymbol{q}) \rangle = \not< T \,, \tag{12.34 a}$$

$$\Sigma_{\text{ph}} \sim T \,. \tag{12.34 b}$$

Using (12.32, 12.33), one then arrives at the following estimate

$$\mu_{\text{ph}} \sim T^{-3/2} \,. \tag{12.35}$$

In addition to the usual scattering from phonons considered here, the scattering in piezoelectric semiconductors (e.g. III–V and II–VI compounds) may contain substantial contributions from phonons that are associated with a polarization (piezoelectric scattering). Charge carriers can also be scattered from optical phonons of higher energy. In this case the description becomes extremely complicated, since the relaxation time approximation (Sect. 9.4) loses its validity.

A further important source of scattering in semiconductors is *scattering from charged defects* (ionized donors or acceptors). A carrier moving past a point-like charged defect experiences a Coulomb interaction, and the scattering cross section Σ_{def} for this "Rutherford scattering" is

$$\Sigma_{\text{def}} \propto \langle v \rangle^{-4} \,, \tag{12.36}$$

where the thermal average $\langle v \rangle \propto \sqrt{T}$ is assumed for the velocity. Equation (12.36) follows from a classical or quantum mechanical treatment of the scattering process (see introductory textbooks on mechanics or quantum mechanics). Because the total scattering probability must also be proportional to the concentration N_{def} of impurities, it follows from (12.32, 12.33, 12.36) that

$$\frac{1}{\tau_{\text{def}}} \propto N_{\text{def}} / T^{3/2} \,, \tag{12.37}$$

and for the mobility, provided the scattering is due only to charged defects, one thus has

$$\mu_{\text{def}} \propto T^{3/2} \,. \tag{12.38}$$

The reciprocal of the total mobility for scattering from defects and phonons is given by the sum of the reciprocal mobilities in (12.35) and (12.38), i.e. the qualitative behavior shown in Fig. 12.12 results. Figure 12.13 shows the experimentally measured dependence $\mu(T)$ for the electron mobility in *n-Ge*, as determined from Hall effect and conductivity measurements; for the purest crystals ($N_{\text{D}} \simeq 10^{13}\,\text{cm}^{-3}$), $\mu(T)$ approaches the theoretically expected dependence for pure phonon scattering (12.35). With increasing donor concentration N_{D}, the additional contribution from impurities (12.38) becomes evident (see also Fig. 12.12).

Figure 12.14 shows that the characteristic trend of the mobility in Fig. 12.12 also manifests itself in the temperature dependence of the conductivity in the region of carrier saturation, where $n(T)$ is approximately constant (Figs. 12.10, 12.11). In this region $\sigma(T)$ shows a maximum, while in the regions of intrinsic conductivity (high T) and in the freeze-out range

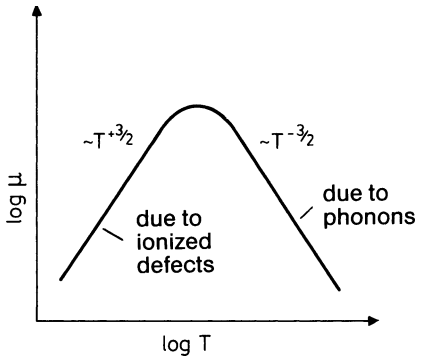

Fig. 12.12. Schematic temperature dependence of the mobility μ for a semiconductor in which scattering from phonons and charged impurities occurs

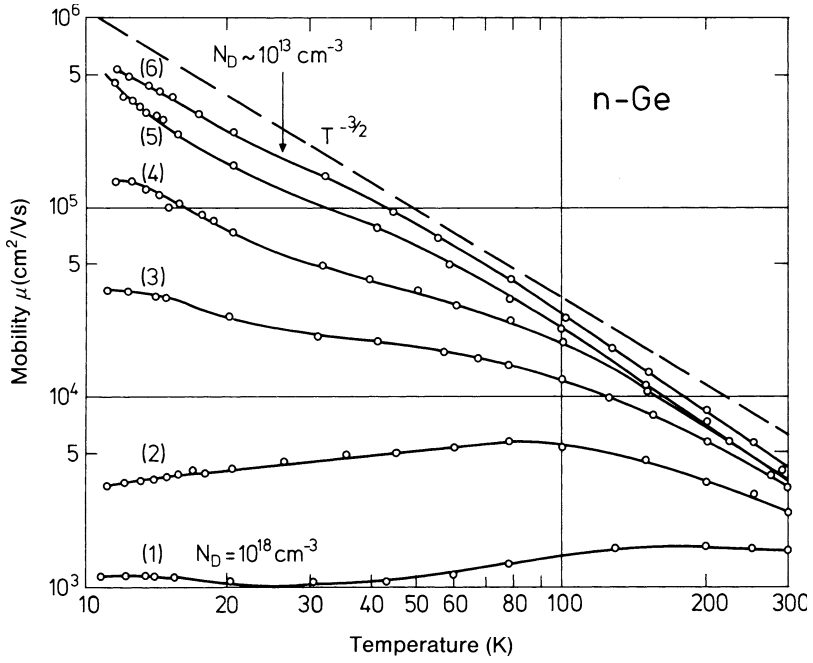

Fig. 12.13. Experimentally determined temperature dependence of the mobility μ of free electrons. For the samples (*1*) to (*6*), the donor concentration N_D varies between 10^{18} and 10^{13} cm^{-3}. The samples are the same as those used for the measurements in Fig. 12.11. (After [12.3])

(low T), the exponential dependence of the carrier concentration $n(T)$ (Fig. 12.11) dominates the weak temperature dependence of the mobility. The experimental results of Figs. 12.11, 12.13 and 12.14 were all obtained from the same Ge samples [numbered from (1) to (6)], so that the conductivity σ can be calculated directly from $n(T)$ and $\mu(T)$.

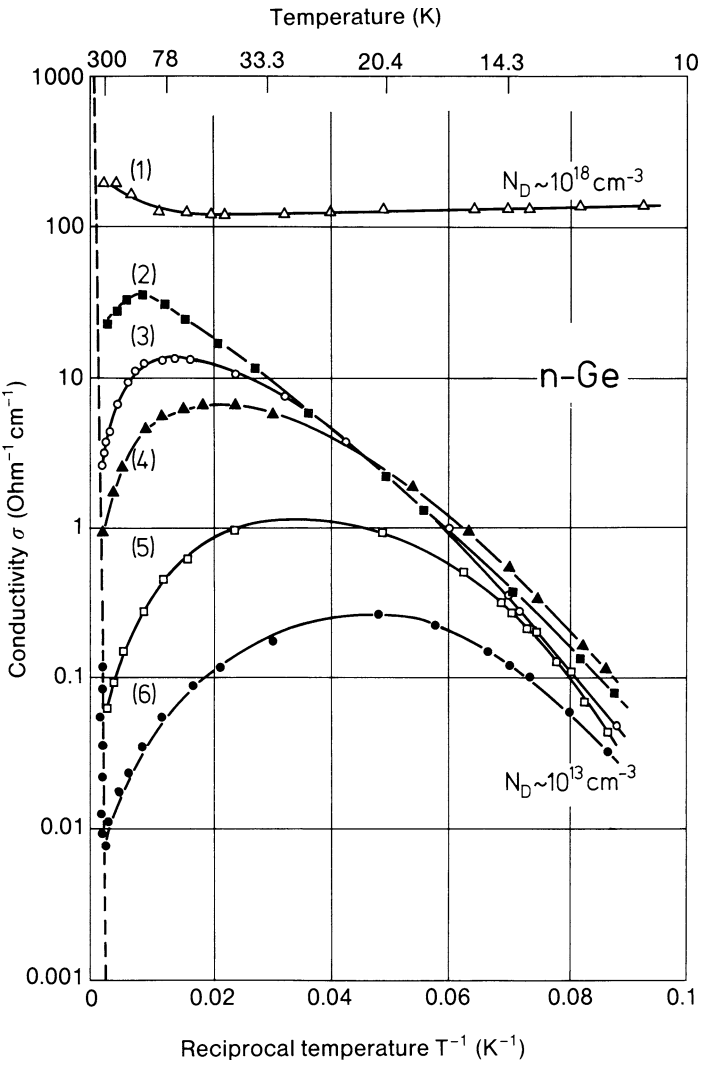

Fig. 12.14. Experimentally measured conductivity σ of *n*-type germanium as a function of temperature. For the samples (*1*) to (*6*), which were also used for the measurements in Figs. 12.11 and 12.13, the donor concentration N_D varies between 10^{18} and 10^{13} cm^{-3}. (After [12.3])

So far the discussion of the conductivity of semiconductors was concerned merely with the ohmic behavior at relatively low electric fields $\mathscr{E}$ where the carrier drift velocity v_D is proportional to the electric field and the mobility $\mu = v_D/\mathscr{E}$ is a constant. In modern semiconductor devices with dimensions in the sub-micrometer range electric fields are frequently in excess of 10^5 V/cm and Ohm's law is no longer valid since the average drift

velocity v_D is no longer proportional to the field strength. According to experiment and theoretical calculations, the proportionality $v_D \propto \mathscr{E}$ remains valid up to fields of about 2×10^3 V/cm for the important semiconductors Si, Ge, GaAs, etc. (Fig. 12.15). For higher fields v_D eventually saturates at a velocity $v_S \approx 10^7$ cm/s for the indirect semiconductor Si, and similarly for Ge. The energy that is continuously transferred to the carriers by the electric field is lost essentially via phonon scattering processes and is thus converted to heat. Scattering due to optical phonons is particularly efficient in this energy range. The carriers are accelerated in the external electric field along the energy band profiles $E(\boldsymbol{k})$, until they reach the energy (referred to E_F) of optical phonons with high density of states (about 60 meV in Si, 36 meV in GaAs). These phonons are then efficiently excited and any further energy gain of the carriers is immediately lost to phonons; the drift velocity saturates as a result. The optical phonons excited in this process are thermalized to low-energy acoustic phonons.

A peculiar behavior is observed in direct semiconductors such as GaAs, InP, GaN. These semiconductors display a negative differential conductivity $\sigma = \partial j / \partial \mathscr{E} = en \partial v_D / \partial \mathscr{E} < 0$ for higher electric fields (Fig. 12.15). In the low field regime ($< 10^3$ V/cm) electrons have a high mobility because of the low effective mass of $0.068\, m_0$ (for GaAs) in the Γ-minimum. As the electrons

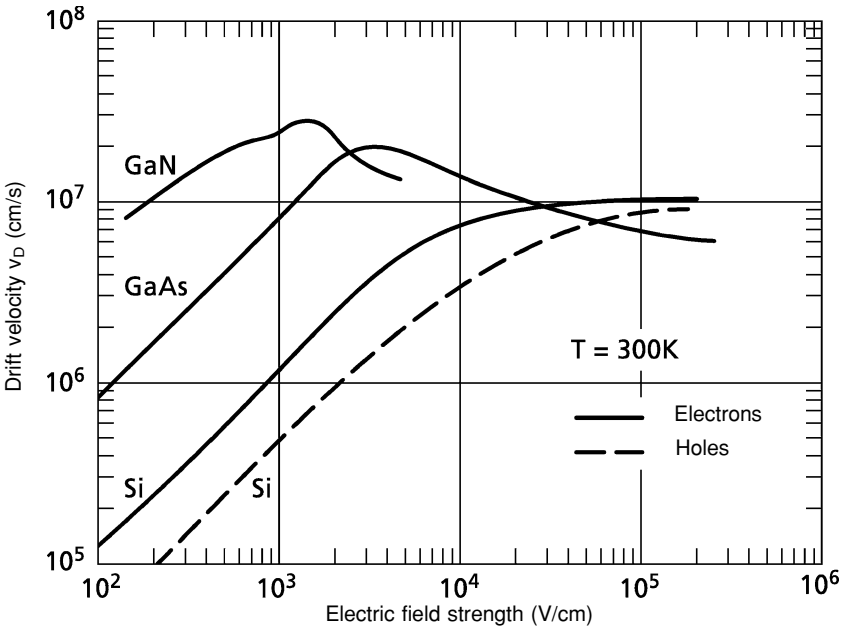

Fig. 12.15. Carrier drift velocity v_D at 300 K as a function of electric field $\mathscr{E}$. The data for Si and GaAs are from a compilation (Sze [12.4]) of experimental results on highly pure, crystalline samples. The curve for GaN is calculated by means of Monte Carlo simulations. (After Gelmont et al. [12.5])

are accelerated to higher kinetic energies effective phonon scattering into the side minima at L and X sets in. Electrons in these minima possess a higher effective mass. Their mobility is therefore lower, and the average drift velocity of the whole ensemble of conduction electrons is thereby reduced. As a consequence, the drift velocity v_D passes through a maximum with increasing field strength and eventually saturates at a lower value. The saturation value is then similar to that of Si. With even further increasing field strength, above 3×10^5 V/cm, the electrons in the side valleys of the band structure with high effective mass are accelerated and the drift velocity v_D would again increase proportional to the field strength albeit with much lower mobility than in the low-field regime ($< 10^3$ V/cm). In the range of extremely high fields, above some 10^5 V/cm this effect is, however, superimposed by the onset of the avalanche breakthrough. The accelerated electrons gain so much energy that they excite more and more electrons from the valence into the conduction band. The conductivity of the semiconductor increases abruptly, avalanche-like, by the multiplication of the number of free carriers. This effect is also used in modern devices.

It is remarkable that the indirect semiconductor Si has a similar saturation velocity at field strengths above 3×10^4 V/cm as the indirect semiconductor GaAs (Fig. 12.15). Hence, the mobility advantage of GaAs with respect to Si at low fields is lost in small devices in which rather high fields occur. A further remarkable aspect of Fig. 12.15 is the high drift velocity of the wide band gap semiconductor GaN. Because of its additional high thermal stability GaN is well suited for applications in the area of fast "high-temperature" devices. Furthermore, the large band gap energies (> 3 eV) of the group III-nitrides in general (GaN, AlN, $Al_xGa_{1-x}N$) enable high breakthrough voltages (up to 100 V) and this class of material is employed in fast high-power devices.

12.6 The p–n Junction and the Metal/Semiconductor Schottky Contact

Modern solid state physics is closely associated with the development of semiconductor devices, i.e. solid state electronics. The operation of almost all semiconductor devices relies on phenomena that are due to inhomogeneities in semiconductors. Inhomogeneous concentrations of donor and acceptor impurities cause particularly interesting conductivity phenomena, which enable the construction of semiconductor devices.

The most important building blocks in semiconductor devices are the p–n junction and the metal/semiconductor contact. In a p–n junction, we have a semiconductor crystal which is p-type on one side, and n-type on the other (Fig. 12.16). In the ideal case (which cannot, of course, be practically realized), the transition from one zone to the other would take the form of a step function.

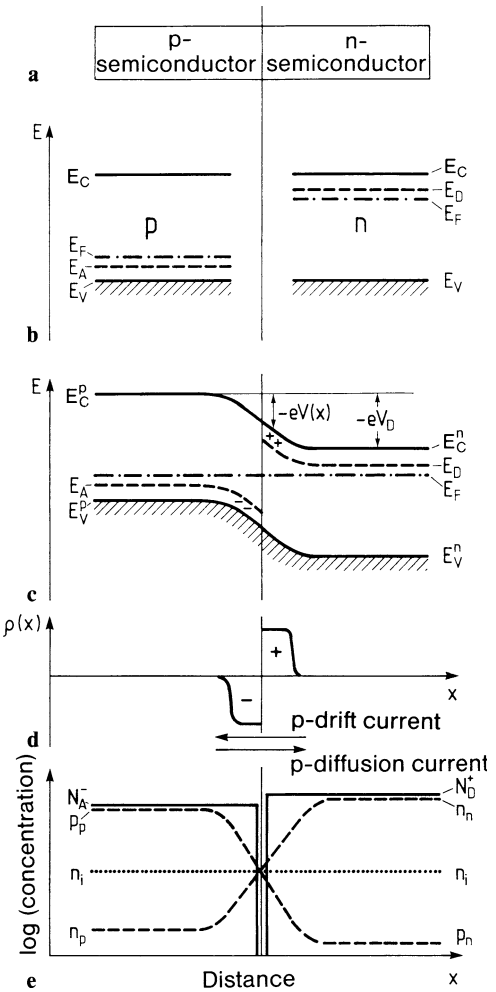

Fig. 12.16 a–e. Schematic representation of a *p–n* junction in thermal equilibrium: **(a)** a semiconductor crystal doped on one side with acceptors (N_A) and on the other side with donors (N_D); **(b)** band scheme for the *n* and *p* sides for the imaginary case of total decoupling of the two sides. E_A and E_D indicate the ground states of the acceptors and donors; E_F is the Fermi level; **(c)** band scheme of the *p–n* junction when the two sides are in thermal equilibrium with one another. The transition from the *p* to the *n* doping is assumed to be abrupt. The position of conduction and valence band edges are denoted E_C^p and E_V^p deep in the *p* region, and E_C^n and E_V^n deep in the *n* region. V_D is the diffusion voltage. In the region of the *p–n* junction, a so-called macropotential $V(x)$ is induced; **(d)** the fixed space charge $\varrho(x)$ in the region of the *p–n* junction due to the ionized impurities; **(e)** qualitative behavior of the concentrations of acceptors N_A, donors N_D, holes *p* and free electrons *n*. The intrinsic carrier concentration is n_i and p_p and p_n denote the hole concentrations deep in the *p* and *n* regions, respectively (and similarly for n_p and n_n). Considered here is the frequently occurring case in which almost all of the donors and acceptors in the interior of the crystal are ionized

In practice, doping inhomogeneities can be created using a variety of techniques. For example, acceptor impurities can be diffused into one region while donors are diffused into another. The dopant atoms can also be implanted as ions: the ions of the required dopant elements are "fired" into the semiconducting material with high kinetic energy (using high electric fields).

In the following we will consider the conduction behavior of a *p–n* junction, consisting for example of a Si crystal, which is doped in the left half with acceptors (e.g. B, Al, Ga) and in the right half with donors (e.g. P, As, Sb). We begin by considering this inhomogeneously doped semiconductor in thermal equilibrium, i.e. without an externally applied voltage (Fig. 12.16).

The p–n Junction in Thermal Equilibrium

Let us first of all imagine that the p- and n-type halves of the crystal are isolated from one another (Fig. 12.16 b), such that the Fermi levels in the two regions lie at different points on the common energy scale. In reality, however, we are concerned with one and the same crystal which simply has an abrupt doping junction. The Fermi level, i.e. the electrochemical potential, must therefore be common to both crystal halves at thermal equilibrium. In the transition zone between the n and p regions there must therefore be a so-called band bending, as shown in Fig. 12.16 c. Within the present semiclassical description, the situation in the transition layer is described by a position-dependent *macropotential* $V(x)$, which reflects the bending of the band structure. This description is possible because the potential $V(x)$ changes only slightly over a lattice parameter. According to the Poisson equation, the macropotential $V(x)$ corresponds to a space charge $\varrho(x)$

$$\frac{\partial^2 V(x)}{\partial x^2} = -\frac{\varrho(x)}{\varepsilon\,\varepsilon_0} \,. \tag{12.39}$$

For the limiting case $T \simeq 0$, where, deep inside the crystal, E_F lies either near the acceptor level (p-type region) or near the donor level (n-type region), one can understand the origin of the space charge from the following qualitative argument: The bending of the bands in the transition region has the effect that acceptors in the p region are pushed below the Fermi level, i.e. are occupied by electrons, whereas in the n region donors are lifted above the Fermi level, i.e. are unoccupied and thus positively charged. The space charge that results from (12.39) thus consists of ionized acceptors and donors that are fixed in space and give rise to a charged double layer across the jump in the doping profile. In the case of carrier freeze-out shown in Fig. 12.16 ($T \simeq 300$ K, E_F between the impurity levels and the middle of the forbidden band), the space charge stems from the fact that the charge of the fixed donors and acceptors in the vicinity of the doping jump is no longer compensated by the mobile electrons and holes of the conduction and valence bands. One therefore speaks of a space-charge region where $\varrho(x) \neq 0$. The concentrations of charge carriers and impurities associated with this space-charge region are represented in Fig. 12.16 e. Well outside the space-charge zone, the donors (N_D^+ if charged) or acceptors (N_A^-) are compensated by equally large electron (n_n) or hole (p_p) concentrations. The subscripts n, p indicate whether the electrons and holes are situated in the n or p regions. These carriers correspond to the type of doping in the respective regions and they are denoted *majority carriers*. Because the electrons and holes are "freely" mobile, electrons diffuse into the p region and holes into the n region. There they are called *minority carriers* and their concentrations are denoted by n_p (electrons) and p_n (holes). In thermal equilibrium, the law of mass action ($n_i^2 = np$) must be fulfilled at each point.

For the concentration of majority carriers (electrons in the n region, n_n, or holes in the p region, p_p), it follows from the arguments in Sect. 12.3 that

$$n_n = N_{\text{eff}}^{\text{C}} \exp\left(-\frac{E_{\text{C}}^n - E_{\text{F}}}{\not{k}\, T}\right), \tag{12.40 a}$$

$$p_p = N_{\text{eff}}^{\text{V}} \exp\left(-\frac{E_{\text{F}} - E_{\text{V}}^p}{\not{k}\, T}\right). \tag{12.40 b}$$

Furthermore we have

$$n_i^2 = n_n p_n = N_{\text{eff}}^{\text{V}} N_{\text{eff}}^{\text{C}} \exp\left(-\frac{E_{\text{C}}^n - E_{\text{V}}^n}{\not{k}\, T}\right). \tag{12.41}$$

The *diffusion voltage* V_{D} – the difference between the maximum and minimum of the macropotential $V(x)$ (Fig. 12.16c) – which is built up in thermal equilibrium, is thus related to the carrier density by

$$eV_{\text{D}} = -(E_{\text{V}}^n - E_{\text{V}}^p) = \not{k}\, T \ln \frac{p_p n_n}{n_i^2}. \tag{12.42}$$

At low temperature in the carrier freeze-out regime it is evident that $|eV_{\text{D}}| \sim E_{\text{g}}$ (Fig. 12.16c). Here, E_{V}^n and E_{V}^p are the valence band edges in the n and p regions.

The state of a semiconductor as represented in Fig. 12.16b–e must be understood as a steady state, because the concentration profiles of "free" carriers as in Fig. 12.16e imply diffusion current (electrons diffusing from right to left and holes from left to right). On the other hand, a space charge as in Fig. 12.16d is associated with an electric field $\mathscr{E}(x)$ and therefore with drift currents of electrons and holes. We represent the corresponding (charge) current densities as:

$$j^{\text{diff}} = j_n^{\text{diff}} + j_p^{\text{diff}} = e\left(D_n \frac{\partial n}{\partial x} - D_p \frac{\partial p}{\partial x}\right), \tag{12.43}$$

$$j^{\text{drift}} = j_n^{\text{drift}} + j_p^{\text{drift}} = e(n\mu_n + p\mu_p)\mathscr{E}_x. \tag{12.44}$$

Here D_n and D_p denote the diffusion constants for electrons and holes, respectively. In (dynamic) thermal equilibrium, the currents exactly compensate one another. In the p and n regions electron-hole pairs are continually created due to the finite temperature, and subsequently recombine. The total current density obeys

$$j^{\text{diff}} + j^{\text{drift}} = 0 \tag{12.45}$$

and thus the separate contributions of the electrons and holes must vanish individually, i.e. for electrons it follows from (12.43–12.45) that

$$D_n \frac{\partial n}{\partial x} = n\mu_n \frac{\partial V(x)}{\partial x}, \tag{12.46}$$

where $\mathscr{E}_x = -\partial V/\partial x$ is used.

If, instead of considering, as in (12.40), the space-charge concentration far outside the space-charge layer where E_C^n and E_V^p are constant, we consider it in the space-charge zone itself, then the conduction band edge is of course described by $[E_C^p - eV(x)]$ and the concentration of the electrons is position dependent with

$$n(x) = N_{\text{eff}}^{C} \exp\left(-\frac{E_C^p - eV(x) - E_F}{\not{k} T} \right) , \tag{12.47}$$

from which it follows that

$$\frac{\partial n}{\partial x} = n \frac{e}{\not{k} T} \frac{\partial V}{\partial x} \tag{12.48}$$

or, by substituting into (12.46),

$$D_n = \frac{\not{k} T}{e} \mu_n . \tag{12.49}$$

This so-called *Einstein relation* between carrier diffusion constant and mobility is valid whenever diffusion currents and drift currents are carried by one and the same type of carrier. A relation analogous to (12.49) also applies of course for holes, since the total hole current must vanish too.

A rigorous treatment of the p–n junction is not simple because, in the Poisson equation (12.39), the exact form of the space-charge density $\varrho(x)$ depends on the interplay of diffusion and drift currents [which in turn depends on $V(x)$]. For the "abrupt" p–n boundary treated here, the following approximate solutions can be given; they are known as the *Schottky model of the space-charge zone*. If we imagine that the zero of the x-axis in Fig. 12.16 is at the junction between the n and the p regions, where the donor (N_D) and the acceptor (N_A) concentrations abruptly meet, then for the space charge we have the general relations

$$\varrho(x > 0) = e(N_D^+ - n + p) \qquad \text{in the } n\text{-region} , \tag{12.50 a}$$

$$\varrho(x < 0) = -e(N_A^- + n - p) \qquad \text{in the } p\text{-region} . \tag{12.50 b}$$

The position-dependent concentrations $n(x)$ and $p(x)$ of "free" charge carriers adjust themselves of course according to the respective distances of the conduction and valence band edges from the Fermi level (Fig. 12.16 c). Although this distance changes slowly and monotonically over the entire space-charge zone, the Fermi function and therefore the occupation in an energy region $\sim 2 \not{k} T$ (300 K) $\simeq 0.05$ eV, which is small compared to the band gap, changes from approximately zero to its maximum value. If one neglects the so-called transition zone of the Fermi function, then the concentration N_D^+ of charged donors not compensated by free electrons, and the corresponding concentration of charged acceptors N_A^-, can be

approximated by a step function (Fig. 12.17a), i.e. the space-charge density becomes

$$\varrho(x) = \begin{cases} 0 & \text{for} \quad x < -d_p \\ -eN_A & \text{for} \quad -d_p < x < 0 \\ eN_D & \text{for} \quad 0 < x < d_n \\ 0 & \text{for} \quad x > d_n \, . \end{cases} \tag{12.51}$$

With this piecewise constant space-charge density, the Poisson equation, for example for the n region ($0 < x < d_n$)

$$\frac{d^2V(x)}{dx^2} \simeq -\frac{eN_D}{\varepsilon\,\varepsilon_0} \tag{12.52}$$

can be readily integrated. The calculation for the p region is of course completely analogous. For the electric field $\mathscr{E}_x(x)$ and the potential $V(x)$ in the n region of the space charge zone, one obtains (Fig. 12.17b,c)

$$\mathscr{E}_x(x) = -\frac{e}{\varepsilon\,\varepsilon_0} N_D(d_n - x) \tag{12.53}$$

and

$$V(x) = V_n(\infty) - \frac{eN_D}{2\varepsilon\,\varepsilon_0}(d_n - x)^2 \, . \tag{12.54}$$

Outside the "Schottky space-charge zone", the potentials are $V_n(\infty)$ in the n region and $V_p(-\infty)$ in the p region (Fig. 12.17c). One notes the inverted form of the band edges $E_V(x)$ and $E_C(x)$ (Fig. 12.16c) compared with the potential energy $[E(x) = -eV(x)]$. Within the Schottky model, the lengths d_n and d_p give the spatial extent of the space-charge zone in the n and p regions, respectively.

From charge neutrality it follows that

$$N_D d_n = N_A d_p \, , \tag{12.55}$$

and the continuity of $V(x)$ at $x = 0$ demands

$$\frac{e}{2\,\varepsilon\,\varepsilon_0}(N_D d_n^2 + N_A d_p^2) = V_n(\infty) - V_p(-\infty) = V_D \, . \tag{12.56}$$

If the impurity concentrations are known, one can thus calculate the spatial extent of the space-charge layer from the diffusion voltage V_D and the difference between the positions of the band edges in the p and n regions.

From (12.55) and (12.56) it follows that

$$d_n = \left(\frac{2\,\varepsilon\,\varepsilon_0 V_D}{e}\,\frac{N_A/N_D}{N_A + N_D}\right)^{1/2} , \tag{12.57a}$$

$$d_p = \left(\frac{2\,\varepsilon\,\varepsilon_0 V_D}{e}\,\frac{N_D/N_A}{N_A + N_D}\right)^{1/2} . \tag{12.57b}$$

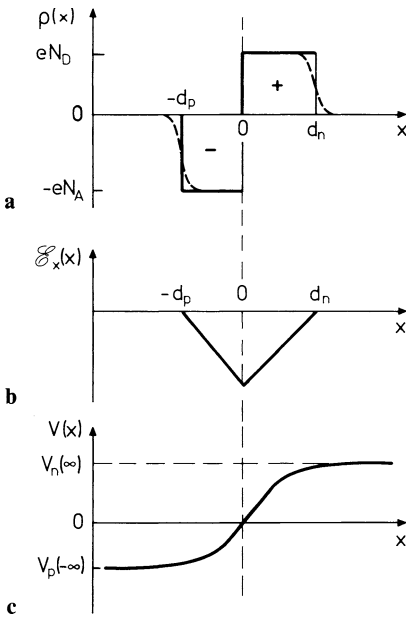

Fig. 12.17 a–c. The Schottky model for the space-charge zone of a *p–n* junction (at $x = 0$). (**a**) Spatial variation of the space-charge density $\varrho(x)$ produced by the ionized acceptors (N_A) and donors (N_D). The real form of the curve (*dashed*) is approximated by the rectangular (*full line*) form; (**b**) behavior of the electric field strength $\mathscr{E}_x(x)$; (**c**) the potential $V(x)$ in the region of the *p–n* junction

Since diffusion potentials are typically of the order of E_g, that is about 1 eV, impurity concentrations of 10^{14}–$10^{18}\,\mathrm{cm}^{-3}$ result in a space-charge zone of size d_n or d_p between 10^4 and 10^2 Å. The electric field strength is therefore between 10^4 and 10^6 V/cm in such space charge zones.

The Biased *p–n* Junction – Rectification

If a time-independent external electrical voltage U is applied to a *p–n* junction, thermal equilibrium is destroyed, and the situation in the *p–n* junction can be described as a stationary state in the vicinity of thermal equilibrium. Because of the depletion of free carriers (depletion zone), the space-charge zone between $-d_p$ and d_n has a considerably higher electrical resistance than the region outside the *p–n* junction. As a result, the potential drop across the space-charge zone accounts for nearly all of the externally applied voltage U. Thus the band scheme of Fig. 12.16 and the form of the potential in Fig. 12.17c do not alter except in the region of the space-charge zone; elsewhere, $E_C(x)$, $E_V(x)$ and $V(x)$ are constant and therefore remain horizontal. The potential drop across the space-charge zone, instead of being equal to the diffusion voltage V_D (at equilibrium: $U = 0$ V), now has the value

$$V_n(\infty) - V_p(-\infty) = V_D - U\,. \tag{12.58}$$

Here U is taken to be positive when the potential on the *p* side is raised relative to the *n* side. The externally applied voltage now influences the size of

the space-charge zone, since the quantity V_D in (12.57) is now replaced by $V_D - U$. One thus has

$$d_n(U) = d_n(U=0)(1 - U/V_D)^{1/2} , \tag{12.59 a}$$

$$d_p(U) = d_p(U=0)(1 - U/V_D)^{1/2} . \tag{12.59 b}$$

In thermal equilibrium, the drift and diffusion currents of electrons are equal and opposite, and the same is true for the hole currents. If an external voltage U is applied, then equilibrium is destroyed. Let us consider, for example, the balance in the electron currents: we are concerned on the one hand with the drift currents of the minority carriers coming from the p region (where electrons are the minority carriers), which are drawn across into the n region by the diffusion voltage V_D. Because these minority carriers are continually generated in the p region by thermal excitation, this current is called the *generation current*, I_n^{gen}. For a sufficiently thin space-charge zone and a sufficiently small recombination rate in this region, each electron coming from the p region that finds its way into the field of the space-charge zone will be drawn from this into the n region. This effect is largely independent of the value of the diffusion voltage and therefore also of the external voltage.

The diffusion current of electrons from the n region, where the electrons are majority carriers, into the p region (called the *recombination current* I_n^{rec}) behaves differently. In this direction the electrons are moving against the potential threshold of the diffusion voltage. The fraction of electrons which can overcome the potential threshold is determined by the Boltzmann factor $\exp[-e(V_D - U)/\not{k}T]$, and is therefore strongly dependent on the externally applied voltage U. The following relations describe the electron currents I_n through a p–n junction with an externally applied voltage U:

$$I_n^{rec}(U=0) \approx I_n^{gen}(U \neq 0) , \tag{12.60}$$

$$I_n^{rec} \propto e^{-e(V_D - U)/\not{k}T} , \tag{12.61}$$

which together yield

$$I_n^{rec} = I_n^{gen} e^{eU/\not{k}T} , \tag{12.62}$$

and therefore a total electron current of

$$I_n = I_n^{rec} - I_n^{gen} = I_n^{gen}(e^{eU/\not{k}T} - 1) . \tag{12.63}$$

The same analysis applies to the hole currents I_p, so that for the characteristic of a p–n junction one obtains

$$I(U) = (I_n^{gen} + I_p^{gen})\left(\exp\frac{eU}{\not{k}T} - 1\right) . \tag{12.64}$$

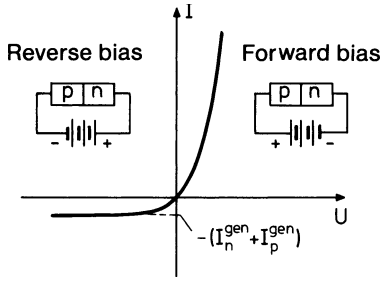

Fig. 12.18. Schematic representation of the current voltage (I–U) characteristic of a p–n junction, together with the corresponding circuit. The maximum current in the reverse direction is given by the sum of the generation currents for electrons and holes

The extremely asymmetric form of this typical rectifying characteristic for the two polarities can be seen in Fig. 12.18.

To derive a quantitative value of the saturation current $-(I_n^{\text{gen}} + I_p^{\text{gen}})$ in the reverse bias direction, a rather more exact treatment of the stationary state in the presence of a dc voltage U is necessary. It follows from the treatment above that the disturbance of thermal equilibrium is due largely to a change in the diffusion currents, while the influence of the external voltage U on the drift currents can, to a good approximation, be neglected. In the framework of the so-called *diffusion current approximation*, it is sufficient to consider only the diffusion currents under the influence of the voltage U.

If we now consider a p–n junction biased in the forward direction (Fig. 12.19), the carrier concentration rises in the space-charge region, as can be seen from Fig. 12.19b. In this situation, the law of mass action $np = n_i^2$ is no longer fulfilled. The Fermi levels well outside the space-charge zone are different by exactly the applied voltage U, corresponding to an energy $-eU$ (Fig. 12.19a). In the space-charge zone, which is no longer in thermal equilibrium, a true Fermi level can no longer be defined. If, as is assumed here, the stationary state is close to thermal equilibrium, then it remains possible to describe the situation approximately with Boltzmann statistics. However, instead of a single Fermi level, two so-called quasi Fermi levels, for electrons (dotted line in Fig. 12.19a) and for holes (dashed line in Fig. 12.19a), must be introduced. Further details of this approach are beyond the scope of this treatment.

Using the approximation that recombination of electrons with holes can be neglected in the space-charge zone, it is sufficient to consider the change in the diffusion current density at the edges of the space-charge zone, i.e. at $x = -d_p$ and $x = d_n$. Here the calculation is particularly easy. We restrict ourselves to the treatment of hole diffusion current densities j_p^{diff}, because the calculation for electrons is analogous.

For the diffusion current at $x = d_n$, it follows from (12.43) that

$$j_p^{\text{diff}}(x = d_n) = -e\,D_p \left.\frac{\partial p}{\partial x}\right|_{x=d_n} . \tag{12.65}$$

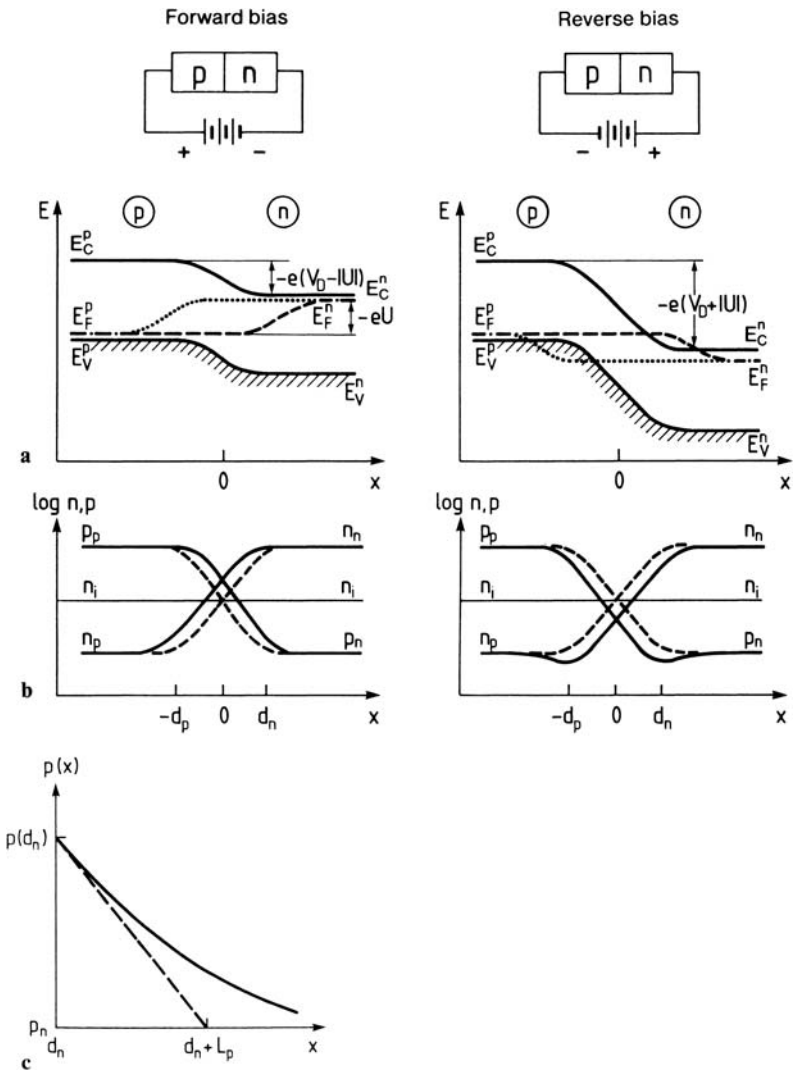

Fig. 12.19 a–c. Forward- and reverse-biased *p–n* junctions (non-equilibrium state). (**a**) Band scheme in the presence of an external voltage $+U$ or $-U$. The Fermi levels E_F^p and E_F^n in the *p* and *n* regions are shifted with respect to one another by eU. In the region of the *p–n* junction, the equilibrium Fermi level (– · –) splits into so-called quasi Fermi levels for electrons (· · ·) and for holes (– – –); (**b**) spatial variation of the concentration of holes *p* and electrons *n* in a biased *p–n* junction (*full line*) and without bias at thermal equilibrium (– – –). The lengths $-d_p$ and d_n give the range of the space charge zone in thermal equilibrium, i.e. without bias voltage. The carrier concentrations deep in the *p* and *n* regions are denoted p_p, n_p and p_n, n_n, respectively; (**c**) spatial variation of the hole concentration $p(x)$ in a forward biased *p–n* junction in the region outside the thermal equilibrium space-charge zone $(x > d_n)$. (Enlarged from Fig. 12.19 b for forward bias)

As will be shown in the following, diffusion theory yields a simple relationship between the concentration gradient $\partial p/\partial x$ and the increase in the hole concentration at $x = d_n$. The resulting hole concentration $p(x = d_n)$ (Fig. 12.19 b) at $x = d_n$, due to the applied bias U, follows from Boltzmann statistics as

$$p(x = d_p) = p_p \exp\left(-e\frac{V_D - U}{\not{k}\, T}\right) , \tag{12.66}$$

$$p(x = d_n) = p_n \exp(e\, U/\not{k}\, T) . \tag{12.67}$$

Here d_n and d_p are the widths of the space-charge zone in thermal equilibrium ($U = 0$ V). The increased hole concentration leads to increased recombination in the n region, which results in a larger electron current. Far into the n region, the current is thus carried by electrons and far into the p region by holes.

The continuity condition implies that the hole concentration in a volume element can only change if holes flow in or out, recombine or are thermally created. The greater the amount by which the hole concentration exceeds the equilibrium concentration p_n, the higher is the recombination rate. With an average lifetime τ_p for a hole, the continuity relation for the non-equilibrium carriers is

$$\frac{\partial p}{\partial t} = -\frac{1}{e}\ \mathrm{div}\, \boldsymbol{j}_p^{\mathrm{diff}} - \frac{p - p_n}{\tau_p} . \tag{12.68}$$

In the stationary case considered here, $\partial p/\partial t$ must be equal to zero and from (12.65, 12.68) we have

$$\frac{\partial p}{\partial t} = D_p \frac{\partial^2 p}{\partial x^2} - \frac{p - p_n}{\tau_p} = 0 , \tag{12.69 a}$$

and thus

$$\frac{\partial^2 p}{\partial x^2} = \frac{1}{D_p \tau_p}(p - p_n) . \tag{12.69 b}$$

Solution of this differential equation yields the diffusion profile

$$p(x) \sim \exp(-x/\sqrt{D_p \tau_p}) . \tag{12.70}$$

Here $L_p = \sqrt{D_p \tau_p}$ is the diffusion length for holes. This is the distance over which the hole concentration decreases from a certain value, e.g. $p(x = d_n)$ in Fig. 12.19 c, by a factor 1/e. From (12.70) and referring to Fig. 12.19 c, we obtain

$$-\left.\frac{\partial p}{\partial x}\right|_{x=d_n} = \frac{p(x = d_n) - p_n}{L_p} , \tag{12.71}$$

where, in the case of an externally applied voltage, we have from (12.67)

$$p(x = d_n) - p_n = p_n[\exp(e\,U/\not{k}\,T) - 1]\ . \tag{12.72}$$

From (12.65, 12.71, 12.72), we finally arrive at the hole diffusion current density due to the externally applied voltage U:

$$j_p^{\text{diff}}\Big|_{x=d_n} = \frac{e\,D_p}{L_p} p_n[\exp(e\,U/\not{k}\,T) - 1]\ . \tag{12.73}$$

Analogous calculations can be carried out for the diffusion current carried by electrons, and the total current density flowing through the p–n junction is simply the sum of these two components. As shown above, it is not necessary to consider the drift currents. Their components from the p and n regions are, within the present approximation, unchanged from their thermal equilibrium values and serve merely to compensate the equilibrium part of the diffusion current, i.e.

$$j(U) = \left(\frac{e\,D_p}{L_p} p_n + \frac{e\,D_n}{L_n} n_p\right)\left(\exp\frac{e\,U}{\not{k}\,T} - 1\right)\ . \tag{12.74}$$

We have thus expressed the generation currents appearing in (12.64) in terms of the diffusion constants and diffusion lengths of electrons and holes, and the minority carrier concentrations p_n and n_p. Figure 12.20 shows an experimentally measured current-voltage characteristic $I(U)$ of a p–n junction. The form of the curve sketched qualitatively in Fig. 12.18 for a rectifier is clearly recognizable.

When an external voltage U is applied to a p–n junction, the spatial extent of the space-charge zone d_n from (12.59), and thus also the charge stored in the space-charge zone, are altered

$$Q_{\text{sc}} \simeq e\,N_{\text{D}} d_n(U) A\ . \tag{12.75}$$

This relation is valid within the framework of the above "Schottky approximation" for the p–n junction, where A is the cross-sectional area of the junction (perpendicular to the current flux) and N_{D} is the concentration of donors as above.

A p–n junction thus has a voltage-dependent "space-charge capacitance" C_{sc}:

$$C_{\text{sc}} = \left|\frac{dQ_{\text{sc}}}{dU}\right| = e\,N_{\text{D}} A \left|\frac{d}{dU} d_n(U)\right|\ . \tag{12.76}$$

From (12.59a) and (12.57a) it follows that

$$\left|\frac{d}{dU} d_n(U)\right| = \frac{1}{2V_{\text{D}}}\left(\frac{2\,\varepsilon\,\varepsilon_0 V_{\text{D}} N_{\text{A}}/N_{\text{D}}}{e(N_{\text{A}} + N_{\text{D}})} \frac{1}{1 - U/V_{\text{D}}}\right)^{1/2}\ , \tag{12.77}$$

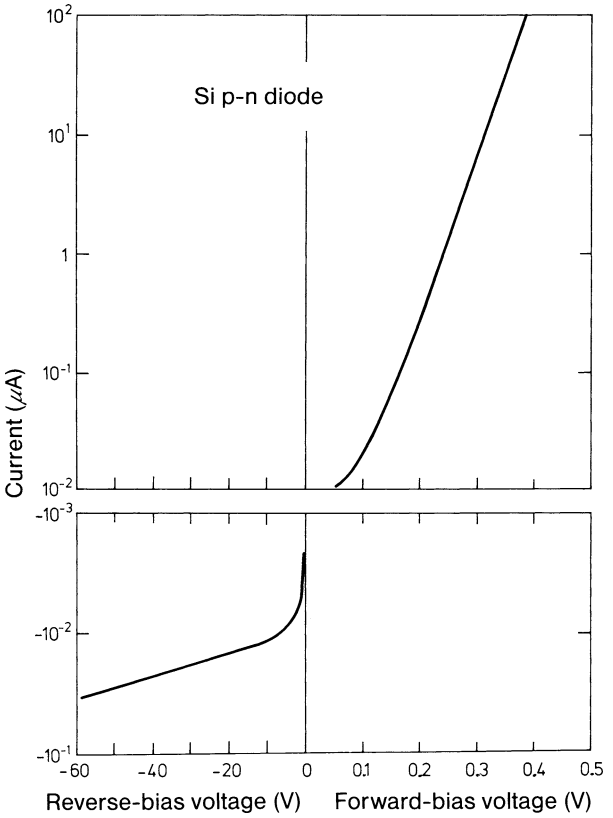

Fig. 12.20. Example of a current-voltage characteristic for a silicon p–n junction. Reverse voltages and currents are given as negative values. (From the advanced lab. course of the II. Physics Institute of the RWTH Aachen)

and so the space-charge capacitance (12.76) can be written

$$C_{\mathrm{sc}} = \frac{A}{2}\left(\frac{N_{\mathrm{A}}N_{\mathrm{D}}}{N_{\mathrm{A}}+N_{\mathrm{D}}}\,\frac{2\,e\,\varepsilon\,\varepsilon_0}{(V_{\mathrm{D}}-U)}\right)^{1/2} . \tag{12.78}$$

The relationship (12.78) implies that

$$C_{\mathrm{sc}}^2 \sim \frac{1}{V_{\mathrm{D}}-U} , \tag{12.79}$$

which explains why a measurement of space-charge capacitance as a function of external voltage is often used to determine the impurity concentrations.

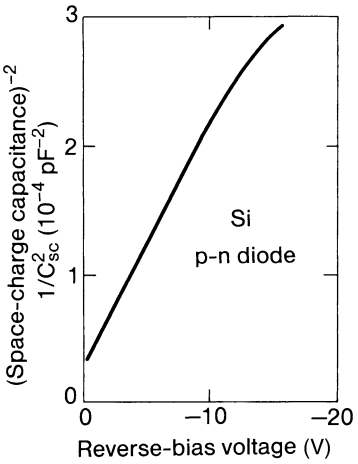

Fig. 12.21. Experimentally determined relationship between the space-charge capacitance and the reverse voltage (indicated by negative values) for the Si *p–n* diode discussed in Fig. 12.19. (From the advanced lab. course of the II. Physics Institute of the RWTH Aachen)

Figure 12.21 shows experimental results for space-charge capacitance as a function of external voltage.

The Metal/Semiconductor Schottky Contact

Because of its rectifying function – similar to the *p–n* junction – the metal/semiconductor contact was used as a device already at the beginning of the 20th century with the advent of microelectronics. When a metal is evaporated onto a clean semiconductor surface under good vacuum conditions, mostly an electronic band scheme as in Fig. 12.22 is established in the case of *n*-doping. The reason is found in the existence of electronic interface states that are formed at the metal/semiconductor interface. Their spatial extension is limited to a few atomic layers around the interface and their energetic distribution is fixed with respect to the conduction and valence band edges of the semiconductor. These interface states, sometimes called MIGS (metal-induced gap states) originate from the Bloch waves in the metal. Bloch waves are extended states in the metal that can not abruptly end at the interface, they must tail into the semiconductor, even in the energy range of the forbidden band, where no electronic bulk states exist on the semiconductor side. In the energy range of the forbidden band these tails of the Bloch states (MIGS) must therefore be mathematically represented by a superposition (Fourier series) of bulk-wave functions of the conduction and valence band, respectively. The energetic band of interface states (MIGS) has thus more conduction-band character (close to the conduction-band edge) or more valence band character (close to the valence-band edge) depending on the percentage of the respective bulk states that constitute the MIGS.

Conduction band states are negatively charged when occupied by electrons and neutral in the unoccupied state (acceptor-like); valence band states, however have a donor-like charging character, i.e. positive when empty and neutral in the occupied state. Thus, a neutrality level exists

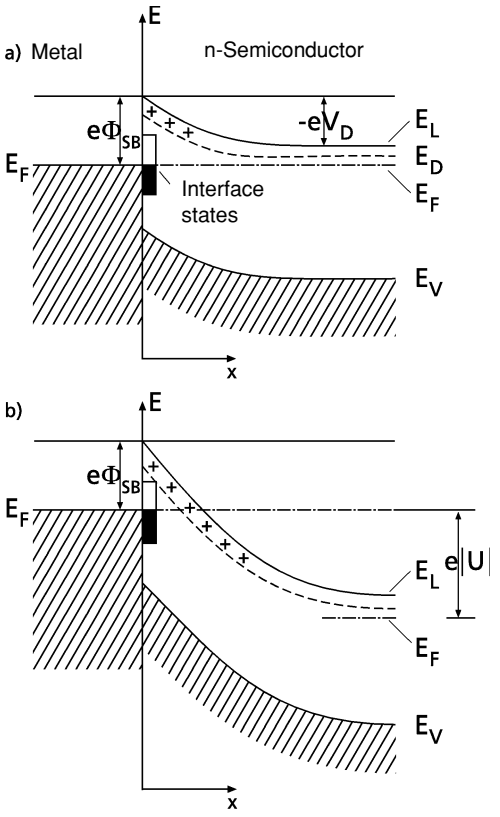

Fig. 12.22. Electronic band scheme of a metal/semiconductor (*n*-doped) junction; pinning of the Fermi-level E_F in interface states near the neutrality level causes the formation of a Schottky-barrier $e\phi_{SB}$ and a depletion space charge layer within the semiconductor. V_D is the "built-in" diffusion voltage. (**a**) In thermal equilibrium, (**b**) under external bias U

within the band of interface states, which separates the more acceptor-like states in the upper part from the more donor-like states in the lower part of the band. When the Fermi energy E_F crosses the band of interface states just at the neutrality level, the interface states as a whole are neutral. Slight deviations of E_F from the neutrality level cause interface charge within the MIGS: negative if E_F lies higher in energy than the neutrality level, positive for E_F below the neutrality level. The band bending within the semiconductor in Fig. 12.22 is thus determined by a charge balance between negative charge Q_{is} in the interface states and the positive space charge Q_{SC} originating from the ionized bulk donor states in the depletion layer.

For atom area densities at the interface of some 10^{14} cm^{-2} electronic interface state densities can reach a similar order of magnitude. Typically, these states are distributed over an energy range of 0.1 to 1 eV within the forbidden band. Interface densities of states per energy can thus easily amount to 10^{15} (eV)$^{-1}$ cm^{-2}. Such a high density of interface states will practically fix (pin) the Fermi level very close to the neutrality level (Fermi-level pinning). Larger deviations of E_F from the neutrality level would cause too high a charge density within the interface that can not be compensated by the space charge

within the semiconductor. Changes of the space charge Q_{SC} by variation of doping or by means of external electric fields thus can shift the Fermi level within the band of interface states with respect to the neutrality level only by amounts of 10^{-2} to 10^{-3} eV. The Schottky barrier (Fig. 12.22) is therefore a quantity that is characteristic for a particular metal-semiconductor junction. Within certain limits it is, of course, slightly affected by atomistic details of the interface structure and morphology.

From a comparison of Fig. 12.22a with Fig. 12.16b it is evident that a Schottky barrier at a metal-semiconductor junction can be described in a simplifying fashion as one half of a *p–n* junction, where the *p*-type semiconductor is replaced by the metal. Obviously, one can interchange the situation; the *n*-side of the *p–n* junction can be replaced by a metal and a Schottky barrier in the *p*-doped semiconductor with hole depletion results. In a *p*-doped semiconductor E_F is close to the valence band edge E_V deep in the bulk (Fig. 12.22). The band curvature in the space charge region of the Schottky contact on *p*-doped material is thus inverse to that of an *n*-type semiconductor. Under the condition of ideal pinning of E_F Schottky barriers for electrons and holes in *n*- and *p*-type material of the same semiconductor sum up to the band gap energy.

As in a *p–n* junction the depletion space-charge region in a Schottky contact exhibits a resistance. An external bias produces a voltage drop essentially across the space-charge zone (Fig. 12.22b). Apart from this spatial range, i.e. within the metal and deep in the semiconductor the voltage drop is negligible and thermal equilibrium is nearly achieved. There, Fermi energies E_F can be defined, the energetic difference of which corresponds to an externally applied voltage. Within the Schottky barrier space-charge region quasi-Fermi levels can be defined similarly as for *p–n* junctions.

The mathematical description of the space-charge region below a metal-semiconductor junction is analogous to a *p–n* junction. The thickness of the Schottky contact space charge region in thermal equilibrium, e.g., is obtained from (12.57a) for the *p–n* junction with the assumption $N_D \ll N_A$ (much higher metallic conductivity as compared with the *p*-type semiconductor) as

$$d = \left(\frac{2\,\varepsilon\,\varepsilon_0 V_D}{e\,N_D} \right)^{1/2} . \tag{12.80}$$

Similarly, the capacity of a metal-semiconductor junction as a function of external bias can be obtained from (12.78) as

$$C = \frac{A}{2} \left(\frac{2\,e\,\varepsilon\,\varepsilon_0 N_D}{V_D - U} \right)^{1/2} , \tag{12.81}$$

where A is the area of the contact.

The rectifying properties of a *p–n* junction are analogously found on a metal-semiconductor Schottky contact, as is evident from Fig. 12.22. Electron transport from the metal into the semiconductor requires that the

carriers overcome the Schottky barrier $e\phi_{SB}$, while in the inverse direction electrons can penetrate the metal from the semiconductor side without any barrier for an external bias $|U| > \phi_{SB}$.

12.7 Semiconductor Heterostructures and Superlattices

Using modern epitaxial methods, such as molecular beam epitaxy (MBE, Panel XVII) or metal organic chemical vapor deposition (MOCVD, Panel XVII), it is today possible to deposit two different semiconductors on one another in a crystalline form. These semiconductors will generally have different electronic properties, and in particular different band gaps. Such layer structures play a particularly important role in devices made from III–V semiconductors such as GaAs, InP, etc. It is also significant that using such epitaxial methods, ternary and quaternary alloys of the type $Al_xGa_{1-x}As$ or $Ga_xIn_{1-x}As_yP_{1-y}$ can be deposited, whose band gaps lie between those of the corresponding binary compounds. By controlled variation of the composition x in $Al_xGa_{1-x}As$, the electronic bandstructure can be continuously adjusted between that of GaAs and that of AlAs. At the composition $x = 0.45$, the alloy changes from a direct-gap semiconductor (like GaAs) to an indirect-gap semiconductor (like AlAs) (Fig. 11.13). Figure 12.23 plots the band gap energies at 300 K for the most important binary and elemental semiconductors against the lattice constant. This plot is of particular interest because, as one might expect, it is possible to

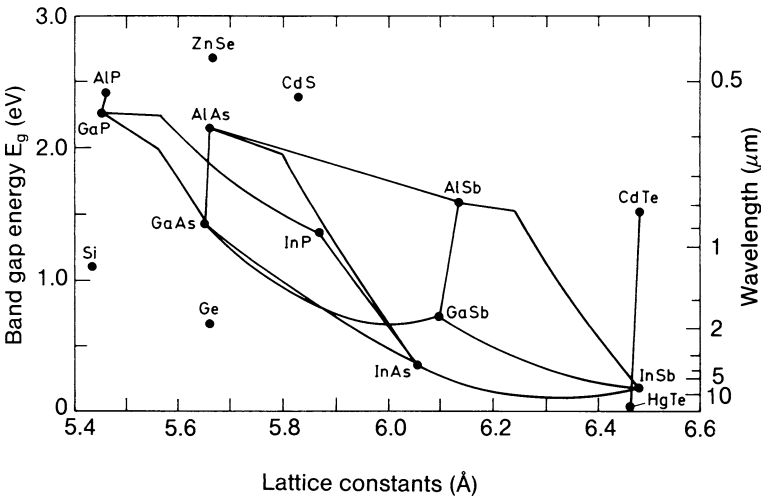

Fig. 12.23. Band gap E_g of some important elemental and binary compound semiconductors plotted against the lattice parameter at 300 K. The right-hand scale gives the light wavelength λ corresponding to the band gap energy. The connecting lines give the energy gaps of the ternary compounds composed of various ratios of the corresponding binary materials

produce particularly good and defect-free epitaxial layers from semiconductors whose lattice constants match. From Fig. 12.23 it can be seen that both band gap and lattice parameter change when one chooses a composition for epitaxy which lies between two binary semiconductors. It can easily be seen that the epitaxy of Ge on GaAs, and vice versa, leads to particularly good lattice matching, and that the alloy system AlGaAs allows a variation of the band gap between 1.4 and 2.2 eV, whereby it is expected that the extremely good lattice matching of the two components GaAs and AlAs should lead to excellent crystalline quality in growing one semiconductor on another. CdTe and HgTe are also an excellent pair of semiconductors which can be grown on one another largely free from defects. The structure consisting of layers of two different semiconductors grown epitaxially on one another is called a semiconductor heterostructure. The width of the transition regions between the two semiconductors can be as little as one atomic layer if a suitable epitaxial method (MBE, MOCVD) is used. In such a heterostructure the band gap changes over distances of atomic dimensions. What does the electronic bandstructure of such a semiconductor heterostructure look like (Fig. 12.24)? Two important questions need to be answered:

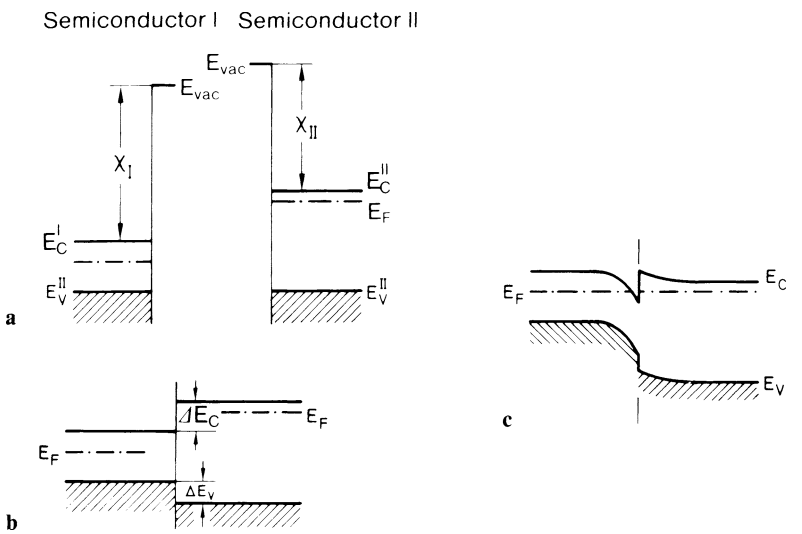

Fig. 12.24 a–c. Band schemes (one-electron energies plotted in real space) for a heterostructure formed from semiconductors I and II. (**a**) Semiconductors I and II are assumed to be isolated; χ_I and χ_{II} are the electron affinities, i.e., the energy between the vacuum energy E_{vac} and the lower conduction band edge E_C. (**b**) Semiconductors I and II are in contact, but not in thermal equilibrium because the Fermi levels E_F on the two sides have not equalized. ΔE_C and ΔE_V are the band discontinuities in the conduction and valence bands, respectively. (**c**) In thermal equilibrium, the Fermi energies E_F in I and II must be identical. Since the band discontinuities ΔE_C and ΔE_V are predetermined, band bending must occur in the two semiconductors

a) How are the valence band edges E_V and conduction band edges E_C to be "lined up"? This question addresses the so-called band discontinuity or band offset ΔE_V (Fig. 12.24b).
b) What band bending occurs in the two semiconductors I and II to the left and right of the junction (Fig. 12.24c)?

In nearly all cases it is possible to treat these two questions separately and independently of one another, because the relevant phenomena involve different energy and length scales. Matching of the two bandstructures takes place within an atomic distance. Interatomic forces and energies are decisive for this process and the electric fields are of the order of the atomic fields ($\gtrsim 10^8$ V/cm). The band bending, on the other hand, takes place over hundreds of Ångstroms, so that in thermal equilibrium the Fermi level has the same value on both sides of the semiconductor interface, and is determined deep inside each semiconductor by the doping level (Fig. 12.24c). As was the case for the *p–n* junction, the decisive condition is that no net current flows in the heterostructure at thermal equilibrium, see (12.45). The band bending in semiconductors I and II is, as in the *p–n* junction, associated with space charge, and the space-charge field strengths are of the order of 10^5 V/cm.

The most important material-related parameters of a semiconductor heterostructure are therefore the valence and conduction band discontinuities, ΔE_V and ΔE_C. The classical, but nowadays revised, assumption was that in the ideal case the two bandstructures of the semiconductors line up such that the vacuum energy levels E_{vac} match one another. This gave a conduction band discontinuity ΔE_C equal to the difference in electron affinities $\Delta\chi$ (Fig. 12.24)

$$\Delta E_C = \chi_I - \chi_{II} = \Delta\chi \,. \tag{12.82}$$

The electron affinity χ, which is the difference between the vacuum energy and the bottom of the conduction band, is a theoretically well-defined quantity in the bulk of a material. It is determined, however, with the aid of surface experiments, in which the bulk value of χ can be strongly altered by surface dipoles (e.g. due to the different atomic coordination at the surface). For the interface between two semiconductors χ cannot be related in a simple way to the bulk quantities. Band offsets are meanwhile well explained by models in which the electronic bands in ideal, abrupt semiconductor heterostructures are lined up so that no atomic dipoles are created, e.g. due to electronic interface states (similar to MIGS, p. 428) or charge transfer in the chemical bonds at the interface [12.6, 12.7]. A detailed theoretical treatment of these models requires a microscopic description of the electronic properties of the few atomic layers at the semiconductor junction. This is beyond the scope of the present book. We therefore treat the band discontinuities ΔE_C and ΔE_V as phenomenological quantities, which may be determined experimentally. A few well-established values are listed in Table 12.7 [12.7]. From a knowledge of the valence band discontinuity ΔE_V and the band gap

Table 12.7. Compilation of a few experimentally determined valence band discontinuities ΔE_V. The semiconductor named first is the substrate, on which the second semiconductor is deposited, generally as a thin strained layer. (After [12.7])

Hetero-structure	Valence band discontinuity ΔE_V [eV]	Hetero-structure	Valence band discontinuity ΔE_V [eV]	Hetero-structure	Valence band discontinuity ΔE_V [eV]
Si-Ge	0.28	InAs-Ge	0.33	CdTe-α-Sn	1.1
AlAs-Ge	0.86	InAs-Si	0.15	ZnSe-Ge	1.40
AlAs-GaAs	0.34	InP-Ge	0.64	ZnSe-Si	1.25
AlSb-GaSb	0.4	InP-Si	0.57	ZnSe-GaAs	1.03
GaAs-Ge	0.49	InSb-Ge	0.0	ZnTe-Ge	0.95
GaAs-Si	0.05	InSb-Si	0.0	ZnTe-Si	0.85
GaAs-InAs	0.17	CdS-Ge	1.75	GaSe-Ge	0.83
GaP-Ge	0.80	CdS-Si	1.55	GaSe-Si	0.74
GaP-Si	0.80	CdSe-Ge	1.30	CuBr-GaAs	0.85
GaSb-Ge	0.20	CdSe-Si	1.20	CuBr-Ge	0.7
GaSb-Si	0.05	CdTe-Ge	0.85		

energies of the two semiconductors, the conduction band discontinuity ΔE_C can of course easily be determined.

The calculation of the space-charge zones is performed in complete analogy to the calculations for a simple *p–n* junction (Sect. 12.6), except that the positive and negative space-charges now occur in two different semiconductor materials I and II with different dielectric constants ε^I and ε^{II}.

The corresponding quantities for a *p–n* heterojunction are depicted in Fig. 12.25. The simplest description is again within the framework of the Schottky model, in which the space charges $-e N_A^I$ and $e N_D^{II}$ are assumed to be constant over their respective space-charge zones d^I and d^{II}. The Poisson equation (12.39, 12.52) can once again be integrated piecewise in semiconductors I and II. It is important to remember that the diffusion voltage V_D is now partitioned between two different semiconductors, i.e. divided into V_D^I and V_D^{II}

$$V_D = V_D^I + V_D^{II} . \tag{12.83}$$

The same is true for an external voltage applied across the *p–n* heterojunction:

$$U = U^I + U^{II} . \tag{12.84}$$

To match the two solutions of the Poisson equation at the interface ($x = 0$), the continuity of the dielectric displacement must be considered

$$\varepsilon_0 \varepsilon^I \mathscr{E}^I(x = 0) = \varepsilon_0 \varepsilon^{II} \mathscr{E}^{II}(x = 0) . \tag{12.85}$$

In analogy to (12.57) the thicknesses of the space-charge zones in semiconductors I and II are given by the formulae

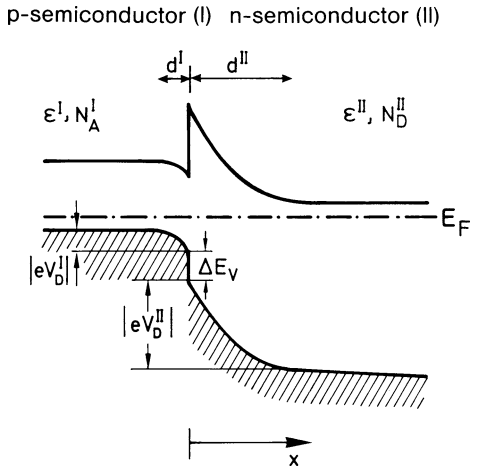

Fig. 12.25. Band scheme of a semiconductor heterojunction; semiconductor I with dielectric function ε^{I} is *p*-doped with an acceptor concentration $N_{\mathrm{A}}^{\mathrm{I}}$, semiconductor II with a dielectric function $\varepsilon^{\mathrm{II}}$ is *n*-doped with a donor concentration $N_{\mathrm{D}}^{\mathrm{II}}$. The space-charge zones have thicknesses d^{I} and d^{II}; the diffusion voltages associated with the space-charge zones are $V_{\mathrm{D}}^{\mathrm{I}}$ and $V_{\mathrm{D}}^{\mathrm{II}}$. ΔE_{V} is the valence band discontinuity. Such a heterojunction consisting of a *p*-type semiconductor with a small band gap and an *n*-type semiconductor with a large band gap is also called a *p–N* heterojunction (small *p* for the small band gap, large *N* for the large band gap)

$$d^{\mathrm{I}} = \left(\frac{2N_{\mathrm{D}}^{\mathrm{II}} \varepsilon_0 \varepsilon^{\mathrm{I}} \varepsilon^{\mathrm{II}} (V_{\mathrm{D}} - U)}{eN_{\mathrm{A}}^{\mathrm{I}} (\varepsilon^{\mathrm{I}} N_{\mathrm{A}}^{\mathrm{I}} + \varepsilon^{\mathrm{II}} N_{\mathrm{D}}^{\mathrm{II}})} \right)^{1/2}, \tag{12.86a}$$

$$d^{\mathrm{II}} = \left(\frac{2N_{\mathrm{A}}^{\mathrm{I}} \varepsilon_0 \varepsilon^{\mathrm{I}} \varepsilon^{\mathrm{II}} (V_{\mathrm{D}} - U)}{eN_{\mathrm{D}}^{\mathrm{II}} (\varepsilon^{\mathrm{I}} N_{\mathrm{A}}^{\mathrm{I}} + \varepsilon^{\mathrm{II}} N_{\mathrm{D}}^{\mathrm{II}})} \right)^{1/2}. \tag{12.86b}$$

Appearing here, in addition to the terms in (12.57), is the externally applied voltage U. The ratio of the voltage drops in the two semiconductors is

$$\frac{V_{\mathrm{D}}^{\mathrm{I}} - U^{\mathrm{I}}}{V_{\mathrm{D}}^{\mathrm{II}} - U^{\mathrm{II}}} = \frac{N_{\mathrm{D}}^{\mathrm{II}} \varepsilon^{\mathrm{II}}}{N_{\mathrm{A}}^{\mathrm{I}} \varepsilon^{\mathrm{I}}}. \tag{12.87}$$

For the case of a simple *p–n* junction ($\varepsilon^{\mathrm{I}} = \varepsilon^{\mathrm{II}}$) in thermal equilibrium ($U = 0$), equation (12.86) reduces of course to (12.57).

Of particular interest are heterojunctions between two different semiconductors with the same doping, so-called isotypic heterojunctions (e.g. *n–N* in Fig. 12.24c). In this case, because of the continuity conditions for the Fermi level, an accumulation space-charge zone for electrons is created on the side of the semiconductor with a smaller forbidden gap, which leads to an extremely large increase in local electron concentration. This is true even when this side of the heterostructure is only very weakly doped, i.e. almost intrinsic (Fig. 12.26a). The high concentration of free electrons in this space-charge zone (semiconductor II) is compensated by a depletion space-charge zone in semiconductor I. This has a strong positive space-charge as a result of the high concentration of ionized donors. These donors have given up their valence electrons to the energetically more favorable potential well in the accumulation zone of semiconductor II. In this way the high density of free electrons is spatially separated from the ionized impurities from which

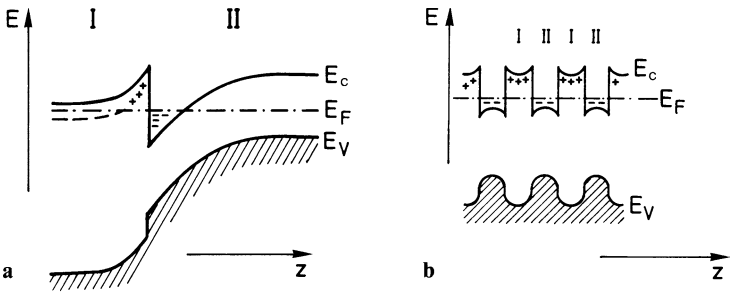

Fig. 12.26. (**a**) Modulation-doped heterojunction consisting of a heavily *n*-doped semiconductor *I* with a large band gap and a weakly (or nearly intrinsic) *n*-doped semiconductor II with a small band gap (*N–n* heterojunction); (**b**) band scheme of a modulation-doped composition superlattice; the layers of the semiconductor I are in each case highly *n*-doped, while the layers of type II are weakly doped or nearly intrinsic

they originate. Impurity scattering, which is an important source of electrical resistance at low temperature, is therefore strongly reduced for this free electron gas. In a homogeneously doped semiconductor, an increase of the carrier concentration requires a simultaneous increase of the doping level, thus leading to increased impurity scattering, which in turn reduces the conductivity. This necessary correlation of higher impurity scattering with increasing carrier concentration does not occur for heterostructures such as that shown in Fig. 12.26. This type of "one-sided" doping in a heterostructure is called "modulation doping".

Electron mobilities of a modulation-doped $Al_xGa_{1-x}As/GaAs$ heterostructure are shown in Fig. 12.27. If the electron mobility were governed by scattering from ionized impurities (ID) and phonons (P) in the same way as in homogeneously doped GaAs, then there would be a characteristic limit resulting from these processes, decreasing at higher and lower temperature. A maximum in the mobility of about $4 \times 10^3\,cm^2/Vs$ would be obtained for a donor concentration of N_D of $10^{17}\,cm^{-3}$ at about 150 K. The experimentally determined mobilities of modulation-doped structures (shaped region) show, however, no reduction in the mobility at low temperature. The variation of the mobility in the shaded region is due to differences in the perfection of the heterojunction achieved by different authors. At temperatures below 10 K, extremely high mobilities of up to $2 \times 10^6\,cm^2/Vs$ are achieved commonly. Best state-of-heart values are around $10^7\,cm^2/Vs$. The increase in mobility can be further enhanced if an undoped $Al_xGa_{1-x}As$ layer with a thickness of about 100 Å is placed between the highly *n*-doped AlGaAs and the weakly doped GaAs during epitaxy. In this way, scattering processes at impurities in the immediate vicinity of the semiconductor junction are eliminated.

For *n*-doping concentrations in AlGaAs of about $10^{18}\,cm^{-3}$, typical thicknesses of the electron enrichment layer are in the region 50–100 Å. The free electrons are confined here in a narrow triangular potential well in the

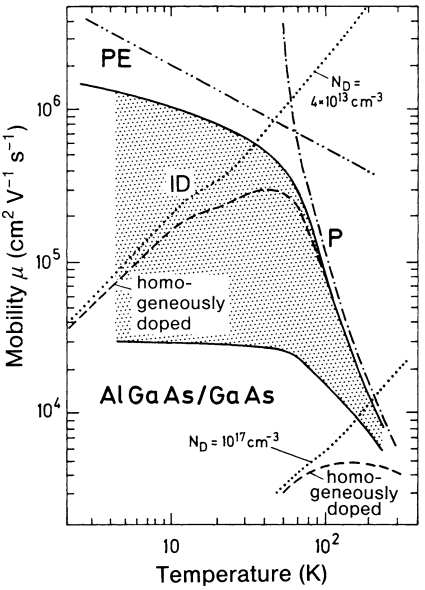

Fig. 12.27. Temperature dependence of the electronic mobility in a quasi 2D electron gas in modulation-doped AlGaAs/GaAs structures; the shaded region represents a great many experimental data. For comparison the mobility curves of homogeneously doped GaAs with donor concentrations $N_D = 4 \times 10^{13}\,cm^{-3}$ and $N_D = 10^{17}\,cm^{-3}$ are also shown (*dashed*). The limiting values of the mobility are determined by the following mechanisms: scattering from ionized impurities (ID) (···); phonon scattering (P) (–·–); piezoelectric scattering (PE) (–· ·–). (After [12.8])

z direction (perpendicular to the heterojunction) (Fig. 12.26a). They can move freely only in the direction parallel to the heterostructure. The wavefunction of such an electron thus has Bloch-wave character only parallel to the heterostructure; perpendicular to it (along z), one expects quantum effects like those for electrons in a potential well (Sect. 6.1). One speaks of a two-dimensional electron gas (2DEG).

Similar effects to those in the simple modulation-doped heterostructure (Fig. 12.26a) appear if one epitaxially grows two heterostructures that are mirror images of one another, or even a whole series of layers of semiconductors I and II with different band gaps (Fig. 12.26b). A series of so-called "quantum wells" are then formed in the bandstructure, in which the free electrons of the conduction band accumulate. Such a structure is shown in Fig. 12.26b; it is called a *composition superlattice*, because the crystal lattice has superimposed on it an artificially created structure of potential wells with larger periodicity. One continues to speak of a modulation-doped superlattice when, as in Fig. 12.26b, only semiconductor I with the larger band gap is heavily *n*-doped, while semiconductor II with the smaller gap is lightly doped. The free electrons in the quantum wells of semiconductor II are again separated from the donor impurities in semiconductor I from which they originate. Impurity scattering is strongly reduced and perpendicular to the superlattice direction (z axis), extremely high mobilities are found at low temperatures as is evident in Fig. 12.27. The band bending shown in Fig. 12.26b, positive in semiconductor I and negative in semiconductor II, corresponds to the sign of the space charge in each region. An unambiguous relationship between these quantities is given

by the Poisson equation (12.39). If the quantum wells in Fig. 12.26 b are sufficiently narrow, i.e. their extension in the z direction is smaller than or on the order of 100 Å, then quantization effects become evident in the z direction for the thin space-charge zone (Fig. 12.26 a).

This so-called z quantization can be described straightforwardly using the time-independent Schrödinger equation for a crystal electron as in Fig. 12.26 a and b. The electron is "trapped" in one direction (along the z axis), while it is free perpendicular to this direction. The potential V is then a function only of z, and with three effective mass components m_x^*, m_y^*, m_z^*, the following Schrödinger equation applies

$$\left[-\frac{\hbar^2}{2}\left(\frac{1}{m_x^*}\frac{\partial^2}{\partial x^2}+\frac{1}{m_y^*}\frac{\partial^2}{\partial y^2}+\frac{1}{m_z^*}\frac{\partial^2}{\partial z^2}\right)-e\,V(z)\right]\psi(\boldsymbol{r})=E(\psi)(\boldsymbol{r})\;. \tag{12.88}$$

With an ansatz of the form

$$\psi(\boldsymbol{r})=\varphi_j(z)\mathrm{e}^{\mathrm{i}k_x x+\mathrm{i}k_y y}=\varphi_j(z)\mathrm{e}^{\mathrm{i}\boldsymbol{k}_\parallel\cdot\boldsymbol{r}}\,, \tag{12.89}$$

equation (12.88) can be separated into two independent differential equations

$$\left[-\frac{\hbar^2}{2m_z^*}\frac{\partial^2}{\partial z^2}-e\,V(z)\right]\varphi_j(z)=\varepsilon_j\varphi_j(z)\;, \tag{12.90}$$

and

$$\left(-\frac{\hbar^2}{2m_x^*}\frac{\partial^2}{\partial x^2}-\frac{\hbar^2}{2m_y^*}\frac{\partial^2}{\partial y^2}\right)\mathrm{e}^{\mathrm{i}k_x x+\mathrm{i}k_y y}=E_{xy}\mathrm{e}^{\mathrm{i}k_x x+\mathrm{i}k_y y}\;. \tag{12.91}$$

The solutions of (12.91) are energy eigenvalues that correspond to the unimpeded motion of an electron perpendicular to z

$$E_{xy}=\frac{\hbar^2}{2m_x^*}k_x^2+\frac{\hbar^2}{2m_y^*}k_y^2=\frac{\hbar^2}{2m_\parallel^*}k_\parallel^2\;. \tag{12.92}$$

For the determination of ε_j in (12.90), the exact form of the potential $V(z)$ must be known. The existence of the space charge means of course that $V(z)$ depends on the density of free electrons and ionized donor cores; i.e., the probability density $|\varphi_j(z)|^2$ is included in the potential via the electron density. Equation (12.90) must thus be solved self-consistently. In a simple approximation, however, $V(z)$ is described by a rigid square-well potential (as in Sect. 6.1) for the case of quantum wells (Fig. 12.26 b), or by a triangular potential for the case of a simple modulation-doped heterostructure (Fig. 12.26 a). If for the square-well potential one assumes in addition infinitely high potential walls, the electron wavefunctions are simple standing waves in the z direction and, as in Sect. 6.1, the eigenvalues ε_j are given by

$$\varepsilon_j\simeq\frac{\hbar^2\pi^2}{2m_z^*}\frac{j^2}{d_z^2}\;,\quad j=1,2,3\ldots\;, \tag{12.93}$$

where d_z is the width of the potential well in the z direction. The total energy eigenvalues for such electron states quantized in the z direction

$$E_j(\boldsymbol{k}_\parallel) = \frac{\hbar^2 k_\parallel^2}{2\,m_\parallel^*} + \varepsilon_j \tag{12.94}$$

are described by a family of discrete energy parabolas along k_x and k_y, so-called subbands (Fig. 12.28 b). These two-dimensional (2 D) subbands have a constant density of states $D(E) = dZ/dE$, as can easily be demonstrated. In analogy to Sect. 6.1 one notes that in the 2 D reciprocal space of wave numbers k_x and k_y, the number of states dZ in a ring of thickness dk and radius k is given by

$$dZ = \frac{2\pi k\,dk}{(2\pi)^2}\ . \tag{12.95}$$

Since $dE = \hbar^2 k\,dk/m_\parallel^*$, one obtains a density of states

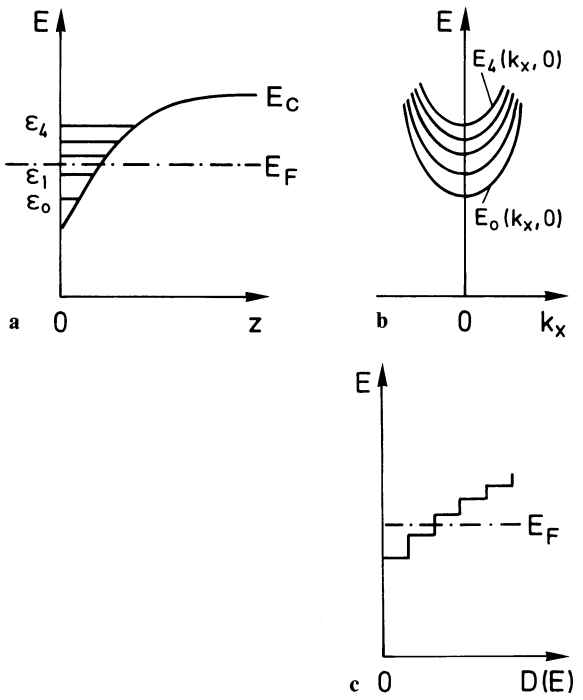

Fig. 12.28 a–c. Quantization of a quasi 2 D electron gas in a triangular potential well, as occurs in the strong accumulation zone at a heterojunction ($z = 0$). (**a**) Behavior of the conduction band edge $E_C(z)$; E_F is the Fermi level. $\varepsilon_1, \varepsilon_2, \ldots$ are the energy levels resulting from the quantization of the one-electron states along z; (**b**) energy parabolas of the subbands, plotted against the wave vector k_x in the plane of free propagation perpendicular to z; (**c**) density of states of the quasi 2 D electron gas reflecting its quantization into subbands

$$D = dZ/dE = m_{\parallel}^{*}/\pi\hbar^2 = \text{const}\,, \tag{12.96}$$

with spin degeneracy accounted for by the factor 2. The total density of states $D(E)$ of all subbands is therefore a superposition of constant contributions, or a staircase function as in Fig. 12.28 c.

Sharp parabolic subbands as in Fig. 12.28 only appear when the eigenvalues ε_j (12.93) are sharp energy levels. This is the case for a single potential well in the conduction band, or when, in a superlattice, the neighboring potential wells are so far apart that the wavefunctions of the individual potential wells do not overlap. If the distance between the potential wells is so small (less than 50–100 Å) that significant overlap between the wavefunctions exists, then this leads to a broadening of the bands. The broadening is completely analogous to that of the individual atomic energy levels of atoms in a crystal (Sect. 7.3). Figure 12.29 shows the theoretically expected broadening for subband energies $\varepsilon_1, \varepsilon_2, \ldots$ for the case of a rectangular superlattice, in which the width of the potential wells d_z corresponds to the spatial separation of two heterojunctions. One sees that the energetically lowest subband ε_1 is noticeably broadened for a periodicity of less than 50 Å, and splits off as a band. For the higher subbands, the broadening begins at even larger distances between the potential wells.

The broadening of the subbands and, in particular, the dependence of the subband energies on the spatial width of the potential wells is clearly seen in photoluminescence experiments. Photoluminescence spectroscopy is

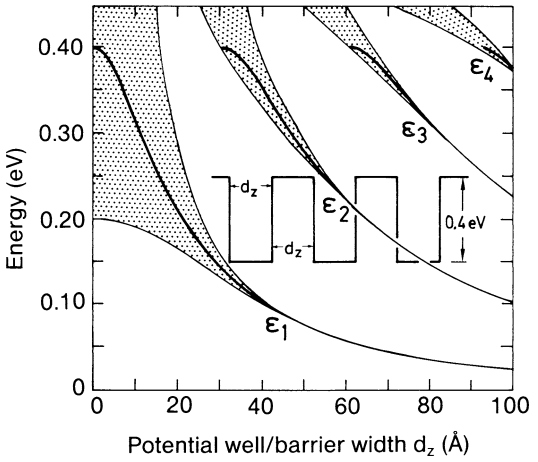

Fig. 12.29. Energy states of electrons confined in the rectangular potential wells (inset) of the conduction bands of a composition superlattice; the potential wells have a width d_z which also corresponds to their distance from one another. For the calculation, an electronic effective mass of $m^* = 0.1\, m_0$ was assumed. The heavy lines in the shaded regions are the results for single potential wells with the corresponding widths d_z; potential wells in a superlattice with sufficiently small separation lead to overlap of wavefunctions and therefore to a broadening into bands (*shaded region*). (After [12.9])

an important optical method for characterizing semiconductor heterostructures and superlattices. It is essentially the reverse of an optical absorption experiment: The semiconductor structure is illuminated with monochromatic laser light of photon energy above the band edge, thus creating electron–hole pairs. These occupy the subbands of the conduction or valence bands of the semiconductor or the corresponding excitonic states (Sect. 11.11). In the case of direct gap semiconductors (Sect. 12.1), the recombination of the free electrons with the holes is associated with strong luminescence, which, for sharply defined subbands, is monochromatic and corresponds, apart from the exciton binding energies (Sect. 11.11), to the energetic separation of the sub-bands for electrons and holes. The experiment has to be performed at low temperatures in order to observe the small energy broadening of the subband structure $(\varepsilon_1, \varepsilon_2, \ldots)$. The low temperature is also the reason why the subband excitons are stable and radiative recombination occurs from the excitonic electron-hole states. Figure 12.30 shows a photoluminescence spectrum taken from an AlGaAs/GaAs multilayer structure at a temperature of 2.7 K. The layer structure consists of four GaAs quantum wells of differing thicknesses (20, 30, 60, 100 Å) embedded between barriers of $Al_{0.3}Ga_{0.7}As$. The four different quantum-well widths give rise to four sharp luminescence lines at different photon energies in the range 1.55–1.75 eV. Since the sub-bands are characterized by the fact that an integral number of half electron wavelengths must more or less fit into the quantum well, narrower quantum wells yield subbands in the valence and conduction bands that lie further apart and thus have higher energy transitions in the photoluminescence. Of course, an exact theoretical analysis of the relationship between luminescence energy and quantum-well width must take into account the Coulomb attraction between the excited electron-hole pairs. This means that the energy difference between electron and hole subbands differs from the energy of the emitted photon by the amount of the exciton binding energy. Since local variations in the width of the quantum well lead to a spread in the spectral position of the emitted photoluminescence line, high

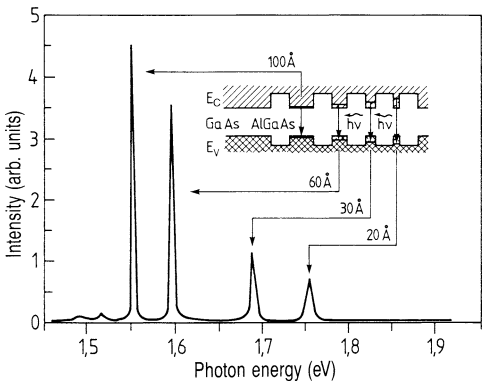

Fig. 12.30. Photoluminescence spectrum of an MBE-grown multiple quantum well structure of $Al_{0.3}Ga_{0.7}As/GaAs$ taken at 2.7 K. The various emission lines result from electron–hole recombination processes in quantum wells of different thicknesses (20–100 Å). The band-structure (E_c, E_v) of the layer structure is indicated showing the correspondence of the quantum wells to the emission lines. (After [12.7])

resolution photoluminescence spectroscopy is particularly well suited to measuring spatial variations in the layer thickness of quantum well layers.

Besides the compositional superstructures considered so far, there is a further type of semiconductor superlattice, the so-called *doping superlattice*. Here the superlattice structure consists of one and the same semiconductor, but the material is periodically and alternately *n* and *p* doped. In principle it is equivalent to a periodic sequence of *p–n* junctions (Sect. 12.6). Because quasi intrinsic (*i* regions) exist between each *n* and *p* zone, these structures also have the name "nipi structures". The production of such lattices is also carried out using epitaxy (MOCVD, MBE, etc., Panel XVII). During the growth process, the *p* and *n* doping sources are switched on alternately.

The interesting properties of nipi superlattices manifest themselves in the position-dependent electronic bandstructure (Fig. 12.31). Because of the periodic sequence of *n*- and *p*-doped regions, the conduction and valence bands must alternately approach the Fermi level. This leads to a periodic modulation of the band edges with position. Excited free electrons (thermal and non-equilibrium) are found in the minima of the conduction band, while excited holes are spatially separated and gather in the maxima of the valence band. This spatial separation of electrons and holes is responsible for the fact that the collision rate between these two particles is drastically reduced. One interesting property which results is an extremely long recombination lifetime for electrons and holes. This is clearly manifest when electron-hole pairs are created by irradiation with light. A photocurrent can exist for considerably longer in such a nipi structure than in a homogeneously doped semiconductor.

Another interesting property of doping superlattices concerns the band gap. In spite of the considerable spatial separation of electron and hole

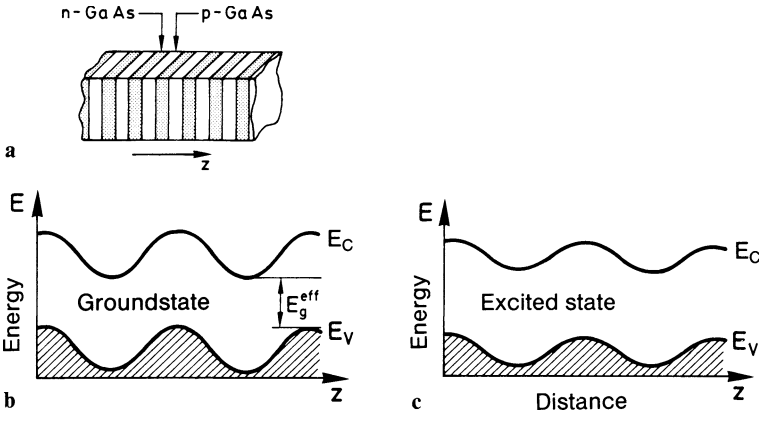

Fig. 12.31 a–c. Schematic bands of a doping (*nipi*) superlattice, in which one and the same semiconductor material (e.g. GaAs) is alternately *n*- and *p*-doped. (**a**) Qualitative sketch of the structure; (**b**) band scheme in thermal equilibrium; (**c**) band scheme under strong excitation of electron–hole pairs to above their thermal equilibrium densities

states, there is nonetheless a certain overlap of the wavefunctions in the transition region (*i* region). This results in the possibility of optical transitions in absorption and emission. As a consequence of the band modulation one thus observes optical transitions at quantum energies below the band edge of the homogeneously doped semiconductor (Fig. 12.32). Nipi structures have an effective band gap E_q^{eff}, which can be adjusted in a controlled way by the respective *n* and *p* doping levels. In the simplest approximation one assumes, as for a *p–n* junction (Sect. 12.6), that all impurities are ionized, and applies the Schottky approximation with a rectangular space-charge distribution. Because of the relationship between band bending and space-charge density [described by the Poisson equation (12.39)], it is immediately clear that a reduction of the space charge results in a decrease of the band bending and therefore a flattening out of the band modulation. The effective band gap becomes larger (Fig. 12.31). The space charge consists of ionized donors (positive) and electrons in the *n* region, and of ionized acceptors (negative) and holes in the *p* region. Sufficiently energetic excitation of electrons and holes, e.g. by irradiation with light, reduces the space charge and also the band modulation. In nipi superlattices, the effective band gap is dependent on the density of optically excited non-equilibrium carriers. Thus the effective band gap can be optically altered. This may be demonstrated in a photoluminescence experiment (Fig. 12.32), in which the emission due to recombination of optically excited electrons and holes is observed as a function of laser excitation power [12.11]. In a GaAs nipi superlattice with a *p* and *n* doping of about $10^{18}\,cm^{-3}$ and a translational period of about 800 Å, the band edge luminescence line at low excitation

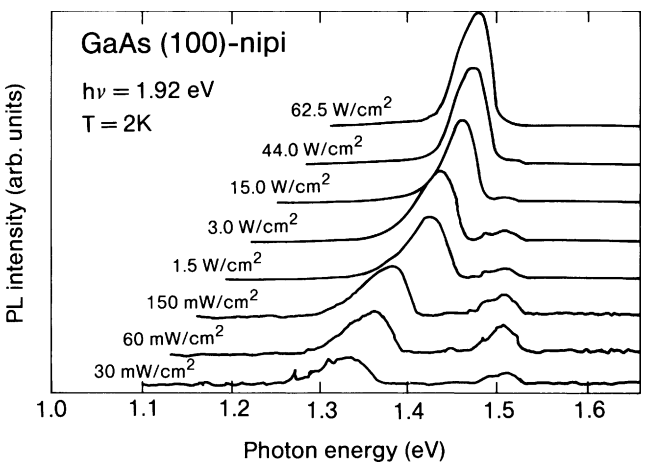

Fig. 12.32. Photoluminescence spectra of a GaAs nipi superlattice, which was deposited by metal organic molecular beam epitaxy (MOMBE) on a GaAs wafer with (100) orientation, *p*-doping due to carbon, *n*-doping due to Si. The spectra were taken with different excitation powers and a photon energy of 1.92 eV at 2 K. (After [12.11])

power appears slightly above 1.3 eV (Fig. 12.32), whereas the band gap in homogeneously doped GaAs at 2 K is about 1.5 eV. With increasing excitation power the luminescence line shifts to a higher energy, approaching the region of the GaAs band edge, as one would expect for a decrease in the band modulation.

Because of their completely new properties which are not present in homogeneous semiconductors, heterostructures and superlattices have opened up a new dimension in the field of microelectronics. Devices based on new concepts are being realized both in the area of very fast circuits and in optoelectronics. This whole domain is a rapidly developing field with great promises for the future [12.12].

12.8 Important Semiconductor Devices

Modern information technology is based on integrated semiconductor circuits (IC) that are used for data processing (logic), data storage (memory) and data transfer in networks, e.g. glass fiber networks, where optoelectronic devices such as detectors and optical emitters (light emitting diodes, LEDs, and lasers) are the fundamental elements. While more than 90% of the semiconductor ICs are fabricated on silicon wafers, the major material base for optoelectronic devices are the III–V semiconductors, and particularly GaAs. The reason is that the elemental semiconductor silicon permits the highest density of integration, i.e. the smallest size of devices, while the much stronger coupling of electronic transitions to the electromagnetic field in the case of direct band gap III–V semiconductors is essential for optoelectronic devices. Of all active devices, transistors are the most important ones for data processing in logic and memory circuits as well as for power electronics. Transistors are three-port devices, i.e. essentially switches, in which an input signal at one contact controls the resistance between two other contacts. Simultaneously, transistors possess amplifying properties for current and/or voltage.

The essential optoelectronic devices, such as detectors, light emitting diodes and lasers, on the other hand, are two-port devices (diodes) that feature a current flow between two contacts only. Devices can be distinguished further by another criterion: Depending on the transport mechanism, be it by only one type of carriers, electrons or holes, or by both types of carriers, devices are called unipolar or bipolar. According to this criterion two classes of transistors exist, so-called bipolar transistors and the field effect transistors (FETs) which have unipolar character. In this sense laser devices and LEDs where the light emission originates from the recombination of electrons and holes are also bipolar devices.

The Bipolar Transistor

The classical bipolar transistor, invented by Bardeen, Brattain and Shockley at the Bell Laboratories (USA) in 1947, consists of two oppositely biased *pn*-junctions (Fig. 12.33a). Accordingly there are both, *npn*- and *pnp*-transistors; in *npn* devices the current is essentially carried by electrons, while in the *pnp*

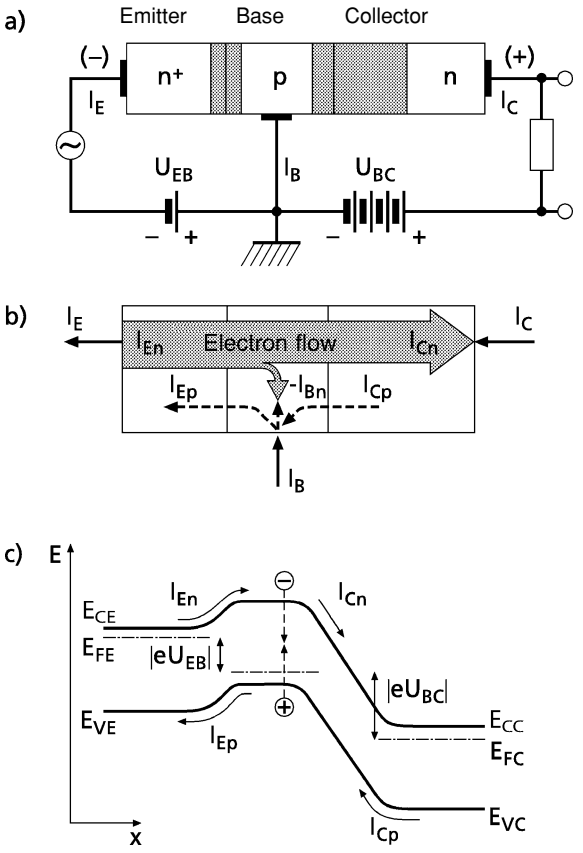

Fig. 12.33. Basic layout and function of an *npn*-bipolar transistor. (**a**) Layout of the transistor together with the external circuitry in the so-called common-base configuration; the space charge regions of the forward-biased emitter (E)-base (B) junction and the reverse-biased base (B)-collector (C) junction are shaded. (**b**) Schematic plot of the currents involved in the function of the transistor. The large shaded arrow depicts the electron current from the emitter to the collector via the base. A small portion of the current is diverted into the base. This current pattern is essential for the performance of the transistor. The directions of the currents (I_E, I_C, I_B, etc.) are drawn according to engineer's convention. (**c**) Electronic band scheme $E(x)$ along the coordinate x from the emitter to the collector, with an emitter-base voltage U_{EB} and a base-collector voltage U_{BC} applied. E_{CE}, E_{VE}, and E_{CC}, E_{VC} are conduction and valence band edges in the emitter (E) and the collector (C) regions, respectively. E_{FE} and E_{FC} are the quasi-Fermi levels in the emitter and collector regions, respectively. The electron-hole recombination process in the base region (which is detrimental to the performance) is indicated by dashed arrows

arrangement holes are the main carrier type. Nevertheless, both carrier types contribute to the device performance, as is expressed by the term "bipolar" transistor. In Fig. 12.33 a the scheme of an *npn*-transistor together with its external circuitry is shown: the first n^+p-junction (high n doping, medium p doping) is biased in the forward direction, such that the p-doped base region is flooded by a high concentration of electrons from the n^+-emitter region through the narrow, low-resistance n^+p space-charge zone. The electrons injected from the emitter into the base region are minority carriers there, and have, therefore, a high tendency to recombine with holes, which represent the majority carriers in the base region. However, in an efficient bipolar transistor the base region is small compared to the recombination length and the major part of the emitted electrons reaches, without recombination the adjacent base-collector junction (biased in the reverse direction). As is seen from the band scheme in Fig. 12.33 c these excess electrons are accelerated by the spatially extended electric space-charge field of the reverse-biased *pn* junction into the collector region. For a sufficiently narrow base region the collector current I_C is smaller than the emitter current I_E but by only a tiny amount, namely by the current I_{Bn} which is lost due to the recombination with holes (majority carriers in the base) (Fig. 12.33 b). The collector current I_C is mainly determined by the forward bias U_{EB} between emitter and base and therefore also by the base current I_B. The collector current I_C is thus essentially controlled by the base current I_B, while the reverse bias at the base-collector junction has only a negligible effect on the current I_C. The external current I_B into the base consists of three contributions: The portion I_{Bn} originates from recombination of holes with electrons being transferred from the emitter region. The hole current (majority carriers in the base) diffuses into the emitter via the valence band barrier. The third contribution is due to holes that are thermally generated within the collector region and are accelerated into the base by the strong space-charge field between base and collector (Fig. 12.33b). For the different current contributions the following relations therefore result:

$$I_E = I_{En} + I_{Ep} \,, \tag{12.97 a}$$

$$I_C = I_{Cn} + I_{Cp} \,, \tag{12.97 b}$$

$$I_B = I_E - I_C = I_{Ep} + I_{Bn} - I_{Cp} \,. \tag{12.97 c}$$

Since the base current I_B is typically below 1% of the collector current I_C the collector current, which flows as an emitter-base current through the base region into the collector, is approximately represented by the relation for a p–n junction (12.74) as

$$I_C \approx I_E \approx A \frac{eD_n n_B}{L_n} [\exp(eU_{EB}/\not{<}T) - 1] \,. \tag{12.98}$$

Here, A is the current-carrying area of the transistor, D_n the diffusion constant of electrons in the base region, n_B their equilibrium concentration

therein, and L_n the electronic diffusion length in the p-doped base. Since the diffusion length L_n is related to the diffusion time τ_n according to $L_n = \sqrt{D_n \tau_n}$, the ratio D_n/L_n can be considered as a diffusion velocity of electrons in the base. The collector and emitter current (12.98) can therefore be interpreted as a product of the number of electrons injected into the base and their diffusion velocity. This diffusion velocity thus determines predominantly the speed of the transistor. The exponential increase of the collector current I_c with base-emitter voltage U_{EB} is plotted in Fig. 12.34 b. This plot is usually called the *transfer characteristics*. The so-called *output characteristics* $I_C(U_{CE})$ are shown in Fig. 12.34 c. They represent essentially the behavior of the reverse-biased base-collector p–n junction, since the base-emitter resistance is small. The output characteristic is therefore the reverse bias currents of a p–n junction, controlled in magnitude by the emitter-base voltage U_{EB}.

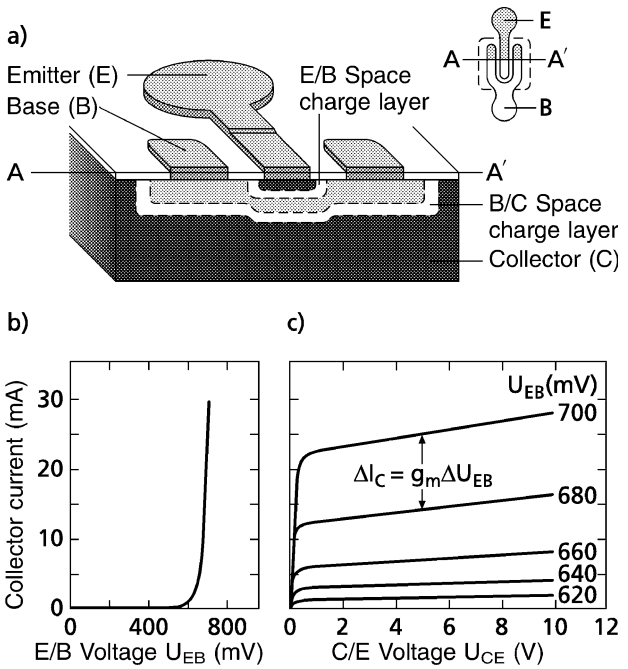

Fig. 12.34. (**a**) Layout of a planar bipolar transistor. The differently doped emitter, base, and collector regions are shaded according to the concentration of dopants, with black denoting the highest concentration. The space-charge layers E/B and B/C are drawn as unshaded. The top view on the metallic emitter (E) and base (B) contact in the inset indicates the position of the cross section through the layer structure along AA'. (**b**) Input characteristic of a bipolar transistor (after Tietze and Schenk [12.13]). (**c**) Output characteristics of a bipolar transistor with emitter-base voltage U_{EB} as a parameter. An important figure of merit of a transistor is the derivative of the collector current I_C with respect to the emitter-base voltage, the *transconductance* $g_m = \partial I_C/\partial U_{EB}$. (After Tietze and Schenk [12.13])

Depending on the different applications, bipolar transistors can be used in connection with three circuitry configurations [12.13]. The common base configuration in which the base is at ground potential was used in the considerations to Fig. 12.33. There, the collector output current I_C is always smaller than the input current I_E by the (small) amount I_{Bn} ($I_{Bn} \ll I_C$). Nevertheless, significant voltage and power amplification can be achieved since the input circuit of the transistor contains the low-resistance emitter-base p–n junction biased in the forward direction, while the output current I_C ($\approx I_E$) flows through the high resistance of the base-collector p–n junction biased in the reverse direction. This permits the use of a high load resistance in the output circuitry with large voltage drop. More frequently used is the emitter configuration where the emitter contact of the transistor is at ground potential. In this configuration, the base current I_B, small in comparison with I_E and I_C, controls the output collector current I_C. The current amplification of this configuration is obvious.

Apart from current amplification the parameters known as *transconductance* $g_m = \partial I_C / \partial U_{EB}$ and *emitter efficiency* $\gamma = \partial I_{En} / \partial I_E$ are the principle figures of merit of a bipolar transistor. The emitter efficiency indicates what percentage of the total emitter current I_E is carried by the majority carriers rather than by holes from the base region. In order to increase the speed of the transistor one has to enhance the conductance of the base region, i.e. the doping level of the base. This, on the other hand, increases the back current of holes I_{Ep} into the emitter and thus causes a decrease of the emitter efficiency γ. A way out of this problem is provided by modern *hetero-bipolar transistors* (HBTs), where the emitter region is formed by a semiconductor layer of a wide band gap material deposited epitaxially on the base-collector system formed by lower band gap materials (e.g., AlGaAs on GaAs). Given a significant valence band offset (Sect. 12.7) this valence band barrier reduces the back-stream of holes from the base into the emitter, despite a large p-doping of the base. Similar concepts are realized by using low band gap semiconductors as the base material in connection with emitter layers possessing a larger band gap.

Field Effect Transistors (FET)

A field effect transistor (FET) is essentially a resistor that is controlled by an external bias voltage. Its architecture is rather simple. A FET consists of a current channel with two contacts, the source and the drain, and a third contact, the gate, which is separated from the current channel by an insulating barrier. The gate contact can therefore be biased with respect to the conducting channel. Depending on the polarity and magnitude of the gate voltage more or less charge is induced in the channel, and the channel conductance is thereby modified. Since only one type of carriers is controlled by the external bias the FET is a unipolar device. There are three main types of FETs (Fig. 12.35), which are distinguished according to the nature of their channels and gate barriers. The most important FET, which is the

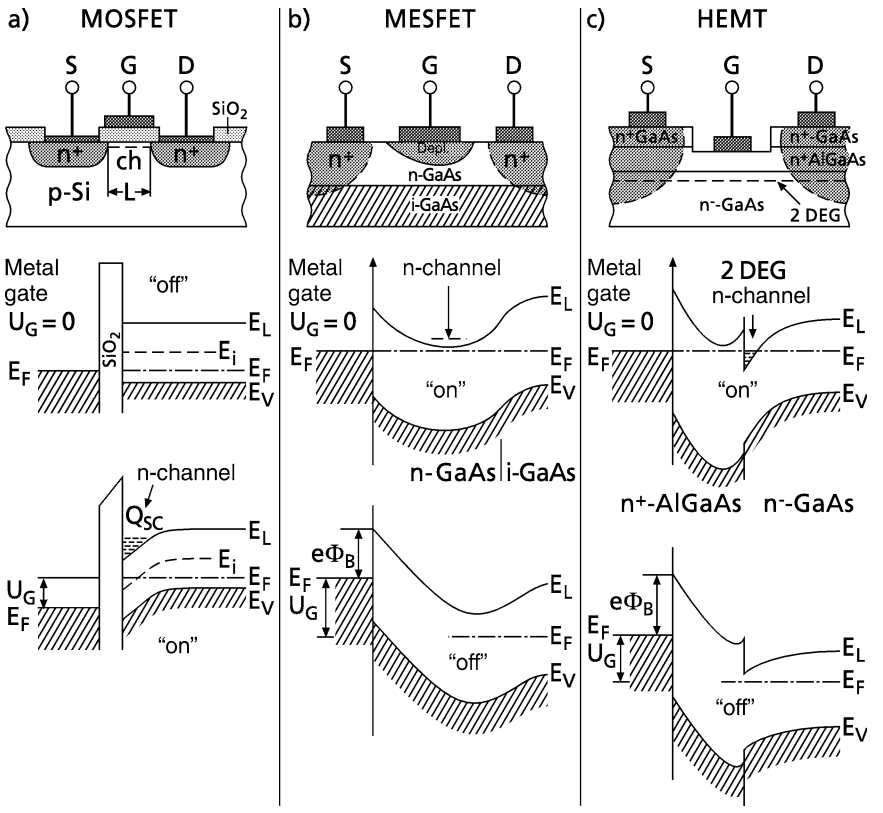

Fig. 12.35. Schematic layout and electronic band schemes for the three most important types of field effect transistors (FET). The figure displays the cross section perpendicular to the wafer surface below the metallic gate electrode. The symbols S, G, and D denote source, gate, and drain, respectively. (**a**) MOSFETs (metal-oxide-semiconductor FET) can only be realized on a Si basis with SiO_2 as gate oxide. The conducting channel and its length are denoted by *ch* and *L*, respectively. For technological reasons the metallic gate electrode is fabricated from highly degenerate doped polycrystalline Si (poly Si). (**b**) MESFETs (metal semiconductor FET) are preferentially realized on III–V semiconductor basis, e.g., GaAs. (**c**) HEMTs (high electron mobility transistor) preferentially realized on III–V semiconductor heterostructures, e.g., AlGaAs/GaAs; the conducting channel is formed by a two-dimensional electron gas (2 DEG) that is connected to the source and drain contacts by highly n^+-doped regions. The band schemes are plotted for the case of zero gate voltage ($U_G = 0$) (upper panel) and for positive (a) and negative (b,c) gate voltages U_G, for MOSFETs and MESFETs and HEMTs, respectively (lower panel); the (quasi-) Fermi level, the conduction band edge, the valence band edge, and the intrinsic energy level are denoted as E_F, E_C, E_V, and E_i, respectively

basis for all highly integrated silicon circuits, be it logic or memory circuits, is the MOSFET (**m**etal **o**xide **s**emiconductor FET). The schematic cross section in Fig. 12.35a displays the essential features: Source and drain regions are formed by highly n^+-doped wells that are generated by ion implantation in a moderately *p*-doped wafer. The two n^+ wells are therefore separated by

a *p*-type region. For any applied voltage between source and drain either one of the two *p–n* junctions is biased in the reverse direction allowing only a small current between source and drain. The *p*-type region between the n^+-source and drain is separated from the metallic gate (*G*) contact by a very thin gate oxide (SiO_2). A bias voltage applied to the gate gives rise to a large electric field across the gate oxide layer that in turn causes a band bending in the underlying semiconductor. For a sufficiently positive gate voltage U_G the initial *p*-region (at $U_G = 0$) below the gate is inverted ($E_i < E_F$) to become an *n*-accumulation layer. The two source and drain n^+-wells are thence connected by a conductive *n*-accumulation channel. The transistor has switched from "off" to "on".

The performance of a MOSFET depends essentially on the perfection of the gate SiO_2 layer and its interface to the underlying Si channel. Any residual trap defect states in the oxide or at the Si/SiO_2 interface would be detrimental to the performance since the charge influenced by the gate voltage would merely reside in the defects rather than affecting the inversion channel. Since only Si is able to form thin, but nevertheless tight and stoichiometrically almost ideal oxides, with atomically smooth interfaces to the Si substrate MOSFETs can be fabricated only on Si wafers. III–V semiconductors such as GaAs, while being superior to Si with regard to electron mobility and thus device speed, do not form perfect oxide overlayers. Even if a high performance of the insulator were granted, the interface between the GaAs substrate and the insulating epilayer exhibits such a high density of interface traps (partially due to non-stoichiometry) that MOSFET structures based on GaAs are very ineffective. The MESFET (**m**etal **s**emiconductor **f**ield **e**ffect **t**ransistor) is therefore the most common FET on III–V compounds (Fig. 12.35 b). A highly resistive, intrinsic (i) or *semi-insulating* GaAs (achieved by doping with Cr) with the Fermi level E_F located near the midgap position serves as the substrate on which a thin, crystalline *n*-doped GaAs layer is epitaxially deposited or generated by ion implantation. Due to its high free electron concentration arising from ionized donors the lower conduction band edge in this epilayer lies close to the Fermi level E_F and a conducting *n*-channel is formed that is separated from the metallic gate by the Schottky barrier below the metal/GaAs junction (Sect. 12.6). Since in this metal/GaAs junction the Fermi level is pinned near midgap at the interface, the electronic bands are bent upwards with respect to the conducting *n*-channel. The metallic gate is thus separated from the *n*-channel by a depletion region of high resistance. In contrast to the *p*-channel, source and drain contacts are highly conductive due to a local high n^+ doping below the metallization layer. Thus, a conductive channel between source and drain exists for $U_G = 0$ in the arrangement of Fig. 12.35 b. This channel can be pinched off by applying a negative voltage U_G to the gate because the negative bias lifts the Fermi level E_F of the metallic gate with respect to E_F in the i-GaAs. Since on the other hand E_F is pinned near midgap ($\phi_B = \text{const}$) at the interface between the gate electrode and the *p*-channel the entire band

scheme is lifted by the applied bias voltage U_G, and the channel becomes depleted of electrons and thus highly resistive. Hence, the transistor switches from "on" to "off" when subjected to a sufficiently high negative gate voltage U_G. The performance depends sensitively on the thickness of the *n*-doped layer as compared with the extension of the Schottky depletion space-charge layer (typically 50 nm) under the gate metal contact.

Complementary to the "normally on"-MESFET described above, the "normally off" MESFET is designed to possess a thin *n*-GaAs channel with a thickness comparable to the extension of the Schottky depletion layer. A sufficiently high positive gate bias bends the electronic bands below the metal gate downwards and opens the channel; the transistor switches to the "on" state only under this positive gate bias.

Since semiconductor heterostructures of high quality can be produced by modern epitaxy techniques such as MBE and MOCVD (panel XVII) the MESFET concept has been extended to a heterostructure FET, the so-called HEMT (**h**igh **e**lectron **m**obility **t**ransistor), which is extremely fast (20–300 GHz) and exhibits very low noise (Fig. 12.35c). Its major fields of application are therefore radar and satellite communication. As a "heterostructure MESFET" the HEMT is switched by a Schottky barrier, separating the metal gate contact electrically from the conducting channel. The conduction channel, however, is a modulation-doped 2DEG at the interface of an AlGaAs/GaAs heterostructure in this case (Sect. 12.7). The high electron concentration of typically $2 \times 10^{12}\,\mathrm{cm}^{-2}$ in the 2DEG originates from donors in the n^+-AlGaAs barrier that are spatially separated from the 2DEG. The conducting 2DEG channel is thus free of ionized scattering centers which results in low noise and high carrier mobility.

A possible modification that exists for all three types of FETs is to invert the type of doping (from *p* to *n*). These "inverted" FETs exhibit the same transistor functions, albeit with inverted external bias (plus to minus). Holes rather than electrons then carry the transistor function. Holes have higher effective masses than electrons, their mobility is lower and therefore *p*-channel FETs are slower than their *n*-channel analogues. Nevertheless, nearly all advanced silicon ICs consist of combinations of *p*- and *n*-channel MOSFETs (CMOS = complementary MOS). The combination of both transistor types allows the realization of logic gates that carry current only within the switching period between two bit-operations. CMOS circuits, therefore, consume less power, which is of significant advantage in the ever-increasing integration density.

FETs show amplification since very small gate currents induce sufficient charge on the gate such that much larger source-drain currents are controlled. Since the resistance of the SiO_2 layer in the case of the MOSFET is much higher than the resistance of the Schottky depletion space-charge layer under the gate of a MESFET or a HEMT, gate currents in MOSFETs are lower than in MESFETs and HEMTs. Characteristic figures of merit for a FET are its transconductance *g* (the ratio between drain current and the applied change

in the gate voltage) and the gate capacity C_g. The ratio between transconductance and gate capacity determines the cut-off frequency $f_{max} = g/C_g$, up beyond which the power amplification drops below 1. In order to improve the high-frequency performance of a FET, its gate capacity should be decreased and the highest possible transconductance should be achieved. Fast, high-performance HEMTs in the material system GaInAs/InP reach transconductances g of about 600 mS/mm and upper cut-off frequencies $f_{max} \approx 300$ GHz at channel lengths of 0.1 μm [12.14].

Typical source-drain current characteristics (I_{SD} vs. U_{SD}) of FETs, in particular of MOSFETs, are shown in Fig. 12.36, with the gate voltage U_G as a parameter. For small U_{SD}, where the inversion space-charge zone along the channel is essentially not affected by a change in U_{SD}, the current I_{SD} is nearly proportional to the applied voltage U_{SD} (quasi-ohmic behavior). Increasing source-drain voltage U_{SD} causes a significant potential drop along the channel within the semiconductor, while the potential on the metallic gate contact remains constant. The voltage drop between the metallic gate contact and the channel (perpendicular to the direction of the channel) thus increases along the channel. This causes a decrease of the channel width (12.80) along the channel, which results in a saturation of I_{SD} for voltages $U_{SD} > 2$ V. The dependence of the saturation current I_{SD} on the gate voltage U_G differs for short- and long-channel transistors. In short channels (Fig. 12.36a), high electric fields of the order of 10^4 to 10^5 V/cm are reached

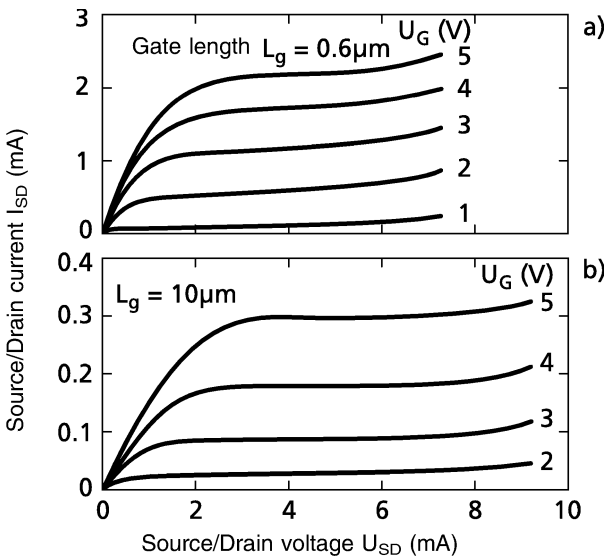

Fig. 12.36. Source-drain current (I_{SD}) vs. voltage (U_{SD}) characteristics of Si MOS-field effect transistors (MOSFETs): $I_{SD}(U_{SD})$ curves are plotted with the gate voltage U_G as a parameter. (**a**) Short-channel MOSFET with gate length $L_g = 0.6$ μm and gate width 5 μm. (**b**) Long-channel MOSFET with gate length $L_g = 10$ μm and gate width 5 μm (after Mühlhoff and McCaughn [12.15])

and the carriers are accelerated up to their saturation velocity v_s (Fig. 12.15). Consequently, the current I_{SD} grows proportionally with the charge induced on the gate and is thus proportional to U_G (Fig. 12.36a). In the case of long channels, one has to consider the fact that the potential difference between gate and channel varies along the channel. In the saturation limit (high source/drain voltage) the gate may be positively biased with respect to the channel near the source, however, still negatively biased with respect to the channel near the drain. As a consequence, the channel is quenched to a high resistance near the drain. A more positive bias on the gate therefore changes not only the carrier concentration locally but also reduces the length where the channel is effectively quenched. In that case a nearly quadratic dependence of the source-drain current I_{SD} on the gate voltage results (Fig. 12.36b). Similar current-voltage characteristics I_{SD} (U_{SD}) as discussed here for MOSFETs are also found for MESFETs and HEMTs.

Semiconductor Lasers

The performance of optoelectronic devices such as photodetectors, solar cells, light emitting diodes (LED) and lasers is based on the interaction of the electromagnetic light field with electronic excitations in the semiconductor. Three different interaction processes exist: (i) an absorption of light quanta by an excitation of electrons from the valence band into the conduction band, or between electronic defect levels within the forbidden band, (ii) an inverse process, the stimulated emission, in which the incident photon stimulates an electron to be de-excited from a state of higher energy into a state of lower energy, and (iii) the spontaneous emission of light. Both, absorption and stimulated emission depend linearly on the strength of the electric field $\mathscr{E}(\omega)$ of the electromagnetic wave (Sect. 11.10). The stimulated emission is coherently coupled with the light field so that the emitted photon is in phase with the stimulating electromagnetic wave. In contrast, spontaneous emission is not coherently coupled to the light field. All three types of interactions are used in photoelectric devices: For photodetectors and solar cells the absorption process is essential. Light emitting diodes (LED) utilize spontaneous emission, while lasers are based on stimulated coherent emission. In all three processes energy conservation is obeyed

$$\hbar\omega = E_2 - E_1 \,, \tag{12.99}$$

with $\hbar\omega$ the photon energy and E_2 and E_1 the energies of the excited and ground state, respectively. Under stationary illumination with light the transition rates must compensate each other. With A_{21} as the probability for spontaneous emission and B_{21} and B_{12} the probabilities for stimulated emission and absorption, respectively, the stationary state is described by

$$\dot{n}_2 = -B_{21}n_2|\mathscr{E}(\omega)|^2 - A_{21}n_2 + B_{12}n_1|\mathscr{E}(\omega)|^2 = -\dot{n}_1 = 0 \,, \tag{12.100}$$

where n_1 and n_2 are the occupation numbers of the ground state and the excited state, respectively. Apart from a pre-factor the coefficients B_{21} and B_{12} are equal to the square modulus of the transition matrix element $|x_{ij}|^2$ for stimulated emission and absorption (11.93) and are therefore equal to each other (11.93). The ratio of the stimulated and spontaneous emission rates is

$$\dot{n}_{2\text{stim}}/\dot{n}_{2\text{spon}} = \frac{B_{21}}{A_{21}}|\mathscr{E}(\omega)|^2 . \tag{12.101}$$

Since laser action requires an excess of stimulated emission processes, a laser is built as a standing-wave resonator for the wavelength to be emitted. Thereby, a high electric field within the lasing medium is generated. Furthermore, the rate of stimulated emission must exceed that of absorption in a laser. Neglecting spontaneous emission because of (12.101) at high light fields $\mathscr{E}(\omega)$ and using the equality $B_{21} = B_{12}$ one obtains from (12.100) the following laser condition:

$$\dot{n}_2 = -(n_2 - n_1)B_{21}|\mathscr{E}(\omega)|^2 < 0 \quad \text{i.e.} \quad n_2 > n_1 . \tag{12.102}$$

This condition describes the *inversion* of the occupation (population) statistics (see Problem 11.8). In thermal equilibrium, just the opposite would hold: the occupation of the excited states is less than the occupation of the ground state. Population inversion is achieved by "pumping" into the excited state. In a semiconductor laser a convenient method for pumping is to bias a *p–n* junction in the forward direction and to flood the space-charge region with non-equilibrium electrons and holes. In writing (12.100–12.102) it was assumed that the final states of the electronic transitions were unoccupied, i.e. their probability to be empty was "one". While this assumption is correct for an ensemble of single atoms in a gas laser, where the occupation of an excited state results from depopulation of the ground state the laser condition has to be modified for the valence band (E_V) and the conduction band states (E_C) of a semiconductor laser. Here, the transition rates are proportional to the density of *occupied initial* and the density of *empty final* states. The condition for population inversion in a *p–n* junction biased in the forward direction thus follows as

$$\begin{aligned} &- B_{21}D_C(E_C)D_V(E_V)f(E_C)[1 - f(E_V)] \\ &+ B_{12}D_V(E_V)D_C(E_C)f(E_V)[1 - f(E_C)] < 0 . \end{aligned} \tag{12.103}$$

In (12.103) the occupied and empty energetic regions of the conduction and valence band, respectively, are represented by sharp energy levels E_C and E_V, the integrals over the corresponding energetic ranges ($\sim kT$) are approximated by constant densities of states $D_C(E_C)$ and $D_V(E_V)$, respectively. $f(E)$ is the Fermi distribution function, which can not be approximated by the Boltzmann distribution at high doping levels. Because $B_{21} = B_{12}$ (12.103) gives:

$$f(E_C)[1 - f(E_V)] > f(E_V)[1 - f(E_C)] \tag{12.104 a}$$

$$f(E_C) > f(E_V) . \tag{12.104 b}$$

The occupation probabilities can be approximated by Fermi functions with E_F^n, E_F^p as quasi-Fermi levels in the n- and p-region, respectively (Fig. 12.37)

$$f(E_C) = [1 + \exp(E_C - E_F^n)/kT]^{-1} , \tag{12.105 a}$$

$$f(E_V) = [1 + \exp(E_V - E_F^p)/kT]^{-1} . \tag{12.105 b}$$

Hence, the condition for population inversion becomes

$$E_F^n - E_F^p > E_C - E_V = E_g . \tag{12.106}$$

For laser operation, the quasi-Fermi levels in the n- and p-doped region must energetically be separated from each other by more than the band gap energy E_g. The p- and n-type regions, therefore, have to be doped deep into degeneracy, in order to reach population inversion by an externally applied forward bias $U = (E_F^n - E_F^p)/e$ (Fig. 12.37 a, b). The current through the p–n junction thus defines the onset of laser action. The emitted light power plotted versus current density through the p–n junction increases slowly due to spontaneous emission, until population inversion is reached at a certain *threshold current density* and laser action starts. At this point stimulated emission becomes the determining mechanism (Fig. 12.38). Threshold

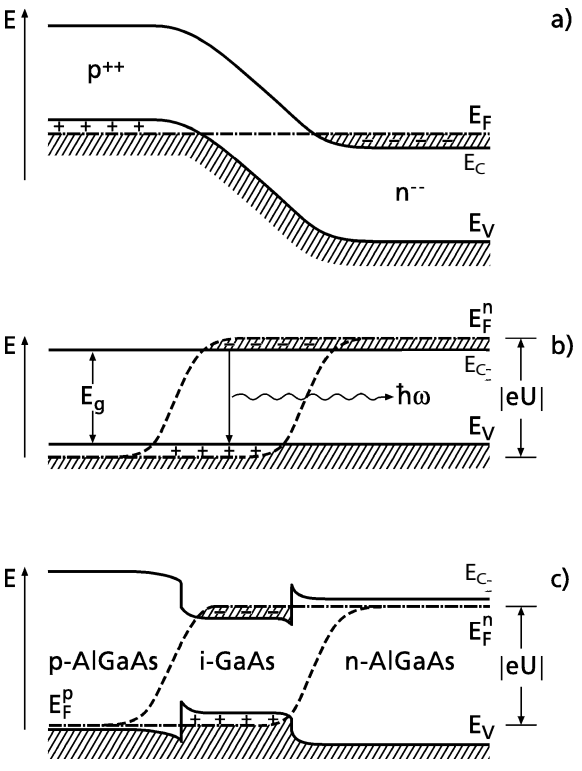

Fig. 12.37. Electronic band schemes $E(x)$ of pn-semiconductor laser structures along a direction x perpendicular to the layer structure: (**a**) Degenerately doped $p^{++}n^{--}$ junction without external bias (thermal equilibrium); (**b**) same $p^{++}n^{--}$ junction with maximum bias U in forward direction; (**c**) double-heterostructure pin junction of p-AlGaAs/i-GaAs/n-AlGaAs with maximum bias U in forward direction. E_F^n, E_F^p are the quasi-Fermi levels in the n- and p-region, respectively; E_C and E_V are conduction and valence band edges

current density is thus the major figure of merit of a semiconductor laser; the lower it is, the more effectively the laser performs. Too high threshold current densities cause too much energy dissipation (Joule heating) which, aside from making the laser less efficient, reduces the lifetime of a laser considerably.

An important breakthrough leading to the commercialization of semiconductor lasers was the invention of the double heterostructure laser. Figure 12.37c shows the electronic band scheme of such a laser, in this case an AlGaAs/GaAs/AlGaAs heterostructure. A lowly doped or intrinsic (i)-GaAs layer is inserted into a *p–n* junction between two *p*- and *n*-doped wide band gap AlGaAs regions. In spite of the low *p*- and *n*-type doping in the AlGaAs regions on both sides the quasi-Fermi levels in the active (i)-GaAs region are located within the conduction and valence band of the GaAs layer under strong forward bias. Population inversion is thus easily achieved without employing highly doped, degenerate semiconductors and the high current densities that accompany the use of degenerate semiconductors. The active *i*-GaAs region is flooded with non-equilibrium electrons and holes, which are confined to the active region by the conduction and valence band discontinuities, respectively. This effect, which causes enhanced light emission is called "*electrical confinement*". Additionally, this structure provides an "*optical confinement*", since for the wavelength of the emitted light (which corresponds to the GaAs band gap) the refractive index of the active GaAs region exceeds that of the adjacent AlGaAs layers (see Problem 11.9). The light originating

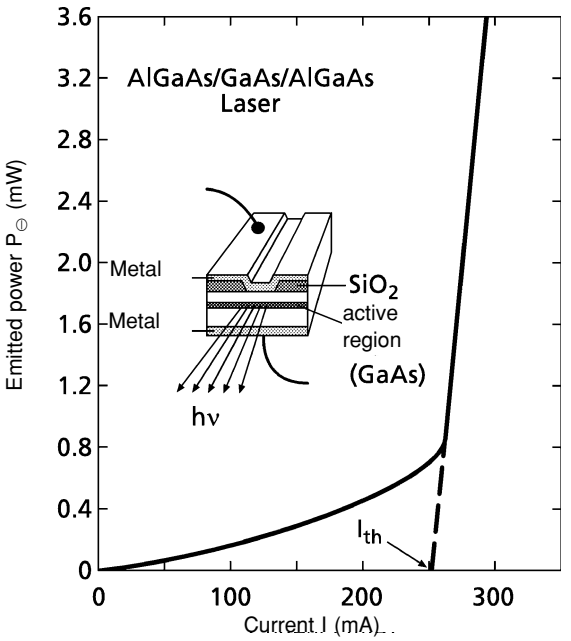

Fig. 12.38. Emission characteristics at room temperature $T = 300\,\mathrm{K}$ of an AlGaAs/GaAs/AlGaAs double-heterostructure laser: the emitted light power is plotted as a function of the current I through the laser; I_{th} is the threshold current where laser action begins. Inset: Schematic layer structure of the laser with emitted radiation ($h\nu$). (After Kressel and Ackley [12.16])

from stimulated emission in the GaAs is therefore totally reflected at the AlGaAs layers; it remains in the active GaAs channel, and is focused and concentrated. The schematic representation of a double heterostructure laser device (inset in Fig. 12.38) shows that the active GaAs channel (resonator) is spatially confined by the two AlGaAs layers on the top and bottom of the GaAs layer, and laterally by a metal stripe deposited into an opening in the SiO_2 mask. The metal contacts the semiconductor layer structure only within this stripe whereby the electric field and the current flow is laterally limited to the width of the stripe.

The optical resonator, i.e. the active GaAs channel (Fabry–Pérot interferometer) requires effective semi-transparent mirrors at its front and back side, through which the light is emitted or reflected. In the simplest case these mirrors are realized by cleavage planes of the GaAs wafer along (110). Typical lengths of these active channels range between 100 and 1000 μm.

The emitted light power as a function of the external current (Fig. 12.38) slowly increases at lower currents due to spontaneous emission. In this current regime the amplification is too low to overcome losses in the resonator. At a certain threshold current I_{th} the amplification of the device matches the resonator losses and laser action starts, and the output light power increases abruptly.

An important figure of merit of a semiconductor laser is thus a threshold current as low as possible (or threshold current density) arising from low losses in the resonator. With increasing temperature the electrical confinement of the laser deteriorates, since carriers populate higher energy states because of a less sharp Fermi edge. The threshold current increases and the laser performs less efficiently. Since Boltzmann statistics are approximately valid, the threshold current I_{th} depends exponentially on the working temperature T:

$$I_{th} \propto \exp(T/T_0) \; . \tag{12.107}$$

A further figure of merit of a semiconductor laser is the critical temperature T_0, which should be as high as possible, so that the laser shows a slow increase of the threshold current with working temperature T. Typical values of T_0 are close to 100 °C for double-heterostructure lasers.

In the double-heterostructure lasers (DHL) considered so far, the vertical extension of the active channel, i.e. the thickness of the embedded GaAs layer is much higher than 200 Å. Carriers in the active zone occupy quantum states of a macroscopic 3D crystal. Further improvement of the laser performance is achieved by incorporating into the active zone even thinner layers (thickness ≤ 100 Å) of a semiconductor with a lower band gap than that of GaAs, e.g., InGaAs. Since $In_xGa_{1-x}As$ is not lattice-matched to GaAs these interlayers are highly strained. They can only be grown without dislocations up to a particular critical thickness. Due to their smaller band gap, i.e. their band offsets with respect to GaAs both in the conduction and valence bands, these thin InGaAs interlayers act as

quantum wells for electrons as well as for holes. The electronic states in the conduction and in the valence band are quantized perpendicular to the layer sequence. The density of states within these layers for each subband is constant and the superposition for all sub-bands is step-like (Fig. 12.28 c). Carriers that occupy the lowest possible energy states are thus even more spatially confined and on even sharper energy levels than in the double heterostructure laser. These so-called *quantum-well lasers* exhibit even better confinement and therefore lower threshold currents and larger critical temperatures T_0.

Problems

12.1 Silicon crystallizes in the diamond structure with a lattice constant $a = 5.43$ Å.

a) How many atoms does the cubic elementary cell contain?
b) Phosphorus atoms incorporated on Si sites act as donors. Calculate the ionization energy of such a donor impurity within the framework of the simple hydrogenic model (dielectric constant $\varepsilon_{Si} = 12$).
c) Calculate the radius r_0 of the first Bohr orbital. What is the probability of finding the electron in its ground state within a sphere of radius $2r_0$? How many Si atoms are located within such a sphere?
d) Discuss the validity of the hydrogenic model and explain its shortcomings.
e) At what impurity concentration does interaction between the donor centers become important? The related impurity band conductivity is assumed to correspond to an average donor separation of $4r_0$.

12.2 The effective mass of electrons at the lower conduction band edge of a semiconductor is three times higher than that of holes at the upper valence band edge. How far is the Fermi energy E_F located from the middle of the forbidden band in the case of intrinsic conduction. The band gap energy E_g must exceed $8 \not< T$; why is this a prerequisite for the calculation?

12.3 A semiconductor with a band gap energy E_g of 1 eV and equal hole and electron effective masses $m_e^* = m_h^* = m_0$ (m_0 is free electron mass) is p-doped with an acceptor concentration of $p = 10^{18}$ cm^{-3}. The acceptor energy level is located 0.2 eV above the valence band edge of the material.

a) Show that intrinsic conduction in this material is negligible at 300 K.
b) Calculate the conductivity σ of the material at room temperature (300 K), given a hole mobility of $\mu_p = 100$ cm^2/Vs at 300 K.
c) Plot the logarithm of the hole concentration, $\ln p$, versus reciprocal temperature $1/T$ for the temperature range 100 to 1000 K.

12.4 Consider the electronic bandstructure of a typical direct-gap III–V semiconductor like GaAs with a high-mobility, sharp conduction band

minimum at Γ and a broad side minimum at somewhat higher energy (cf. Fig. 12.4).

a) Discuss qualitatively the dependence of the average drift velocity of free electrons on the strength of an externally applied electric field taking into account scattering into the side valley of the conduction band.
b) Give a similar discussion of an indirect-gap semiconductor such as Si, where only the broad side minimum of the conduction band is occupied (Fig. 7.13).

12.5 Calculate the Hall coefficient R_H for an intrinsic semiconductor in which electric current is carried by both electrons and holes.

a) Discuss the temperature dependence of the Hall coefficient $R_H(T)$ of the intrinsic semiconductor, when hole and electron mobilities μ_p and μ_n are assumed to be equal.
b) In an experimental measurement of the Hall voltage, reversing the sense of the magnetic field yields a different magnitude for this quantity. Explain the origin of this effect.

12.6 Consider the process of photoconductivity in semiconductors. Concentrations of electrons and holes above those of thermal equilibrium are produced from defect sites or from the valence band by means of photoionization.

Write down the rate equations for the direct recombination of electron–hole pairs for the cases of (a) intrinsic conduction and (b) *n*- or *p*-doping.

Calculate the stationary photocurrent at constant illumination and the time dependence of the photocurrent after the illumination is switched off.

12.7 Explain the operation of a solar cell in which a *p–n* junction is illuminated. Assume a closed circuit containing an external resistance *R*. Make qualitative sketches of the current–voltage relationships with and without incident light. What is the minimum frequency which the light must have? What is the theoretical efficiency as a function of the frequency of the light? How can the current–voltage characteristic be used to obtain the photoelectrically generated power? Discuss the various parameters that influence the efficiency. Why is it advantageous to combine several *p–n* junctions of semiconductors with different band gaps using thin-film technology?

12.8 Small-gap semiconductors such as InAs ($E_g = 0.35$ eV), and InSb ($E_g = 0.18$ eV) usually exhibit surface Fermi level pinning within the conduction band (approximately 100 meV above the lower conduction band edge E_C for InSb). Plot qualitatively the band scheme (band energy versus spatial coordinate) in the vicinity of a metal contact to such a semiconductor that is highly *n*-doped. What is the electrical resistance behavior for both bias directions?

12.9 Calculate the transconductance $g_m = (\partial I_{SD}/\partial U_G)$ as a function of gate capacity C_g and gate length L_g for a field effect transistor (MOSFET, MESFET or HEMT, Sect. 12.8). Note that even without an external gate bias ($U_G = 0$) a "built-in" band bending under the gate with a threshold voltage U_0 is present. For short-channel transistors one can assume in a simple approximation that the carrier velocity in the channel is constant and equal to the saturation velocity v_s for high field strength (Fig. 12.15). The source-drain current I_{SD} consists of the 2-dimensional (2D) carrier density n_{2D} which is induced in the channel by the action of the gate bias U_G.

a) Using Fig. 12.15, estimate the transit time for carriers under the gate for a short channel AlGaAs/GaAs HEMT with gate length $L_g = 100$ nm.
b) What gate capacity can be reached in these transistors for a transconductance $g_m = 150$ mS?

Panel XIV
The Hall Effect

To separately determine the carrier concentration n and the mobility μ appearing in the conductivity $\sigma = n e \mu$, one measures both the conductivity and the Hall effect. A current I is passed through a crystal sample and a magnetic field (magnetic induction $\boldsymbol{B}$) is applied normal to the current. Using two contacts placed opposite to one another (perpendicular to I and $\boldsymbol{B}$), a so-called Hall voltage U_H can be measured (open circuit measurement, $I_H = 0$). This is shown schematically in Fig. XIV.1. Since measured in the absence of current, the Hall voltage U_H which builds up in the sample in the y direction is exactly compensated by the Lorentz force on an electron, which also acts in the y direction. The Lorentz force on an electron moving in the x direction with velocity v_x is:

$$\begin{aligned} F_y &= -e(v \times \boldsymbol{B})_y - e\mathscr{E}_y \\ &= e v_x B - e\mathscr{E}_y = 0 \,. \end{aligned} \tag{XIV.1}$$

Here $\mathscr{E}_y = U_H/b$ is the so-called Hall field. Assuming that the current is carried exclusively by electrons (n-type semiconductor or metal), we have

$$j_x = I/(b\,d) = -n\,e\,v_x \tag{XIV.2}$$

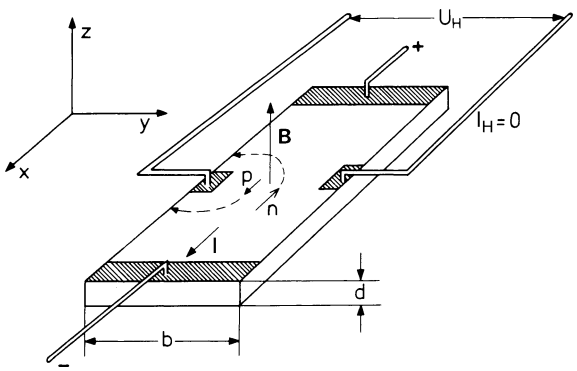

Fig. XIV.1. Schematic set-up for a Hall effect measurement. ***B***: magnetic induction; *i*: current through the sample; U_H: Hall voltage. The dashed lines are the paths which electrons and holes would follow in the absence of a compensating circuit for U_H and with a current flow ($I \neq 0$). For $I_H = 0$, the carriers are forced to move in straight lines (parallel to x) due to the build-up of the Hall voltage U_H

and it thus follows that

$$\mathscr{E}_y = \frac{U_H}{b} = -\frac{1}{n e} j_x B = -\frac{1}{n e}\frac{IB}{bd} . \tag{XIV.3}$$

The quantity $R_H = -(ne)^{-1}$ is called the Hall constant. It can be determined from measurements of U_H, i and B via the relation

$$U_H = R_H I B/d . \tag{XIV.4}$$

The sign of R_H gives the type of carrier (a negative sign corresponding to electrons) and its absolute value gives the carrier concentration n.

If the current in a semiconductor is carried by both electrons (concentration: n, mobility: μ_n) and holes (p, μ_p), then an analogous calculation gives the following expression for the Hall constant (Problem 12.5)

$$R_H = \frac{p\mu_p^2 - n\mu_n^2}{e(p\mu_p + n\mu_n)^2} . \tag{XIV.5}$$

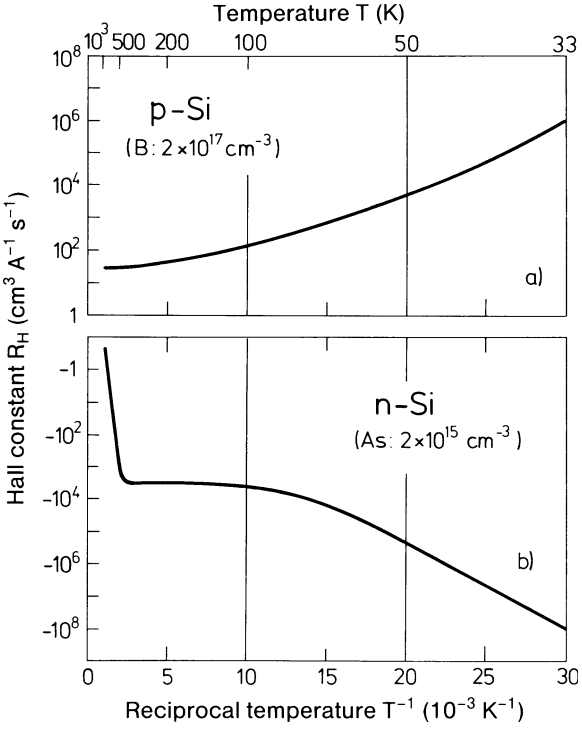

Fig. XIV.2 a, b. Temperature dependence of the Hall constant R_H for p-type (**a**) and n-type silicon (**b**). For p-type with a boron concentration of 2×10^{17} cm^{-3} intrinsic conductivity sets in at temperatures in the vicinity of 1300 K. The curve in part (**a**) would then go through zero and finally convert to the intrinsic branch of (**b**) [XIV.1]

Figure XIV.2 shows experimentally determined values of the Hall constant R_H for boron-doped *p*-silicon (a), and arsenic-doped *n*-silicon (b). Because the Hall constant gives the reciprocal carrier concentration in units of the elementary charge *e*, the shape of the curve in a logarithmic plot is similar (at least in the temperature region 33–500 K) to a typical curve of the carrier density for semiconductors (see e.g. Fig. 12.10). The gradients in the region 33–50 K are governed by the ionization energies of the acceptors and donors (12.27). The steeply rising section of the curve in Fig. XIV.2 b at about 10^3 K reflects the intrinsic conductivity due to electron-hole pair creation. The different signs of R_H in Fig. XIV.2 a, b correspond to the different kinds of carriers in the *p*- and *n*-doped material.

Reference

XIV.1 F.J. Morin, J.P. Maita: Phys. Rev. **96**, 29 (1954)

Panel XV
Cyclotron Resonance in Semiconductors

The effective masses of electrons (m_n^*) and holes (m_p^*) in semiconductors can be determined by cyclotron resonance. A crystal sample is placed in a variable, static magnetic field $\boldsymbol{B}$ and the absorption of a high-frequency alternating field is measured as a function of $\boldsymbol{B}$.

Electrons in the static magnetic field move in $\boldsymbol{k}$-space on surfaces of constant energy in a plane perpendicular to $\boldsymbol{B}$ and around the magnetic field axis (Panel VIII). In real space there is also a closed orbit. For crystal electrons, however, this orbit, like the $\boldsymbol{k}$-space orbit, is not simply a circle (Fig. XV.1). Absorption of the high frequency alternating field occurs when the orbital frequency is exactly equal to the period ω_c ($\omega_c = -e\boldsymbol{B}/m_n^*$). Orbits in $\boldsymbol{k}$-space which enclose an extremal area (maximum or minimum) are especially important since the number of states per frequency interval is particularly high at these points. The high-frequency absorption then has a clear maximum, and for the interpretation of the data one only needs to consider "extremal orbits".

As examples of cyclotron resonance in semiconductors we consider Si and Ge. In a semiconductor, electrons and holes are found in the vicinity of the conduction band minimum and valence band maximum, respectively. As shown in Sect. 12.1, the surfaces of constant energy around the conduction band minimum have the form of ellipsoids with rotational symmetry around either the [100] or [111] directions. For an arbitrary orientation of the magnetic field, there is a different extremal orbit for each ellipsoid pair along a particular [100] or [111] axis. For silicon there are three different values corre-

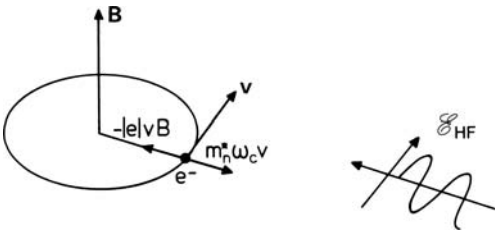

Fig. XV.1. Classical description of cyclotron resonance for electrons with an effective mass m_n^*. On a stable orbit around the magnetic field axis, the centrifugal force and the Lorentz force are in equilibrium. In general, the high frequency field $\mathscr{E}_{HF}$ is maintained at a constant frequency and the magnetic field $\boldsymbol{B}$ is varied

sponding to the three different [100] axes, and for germanium there are four, one for each of the different [111] directions in space (Fig. XV.2 a). If the magnetic field is orientated in a symmetry plane, these numbers are reduced. In Fig. XV.2 b the cyclotron resonance spectra of Dresselhaus et al. [XV.1] are shown. The magnetic field was in the (110) plane. Under these condi-

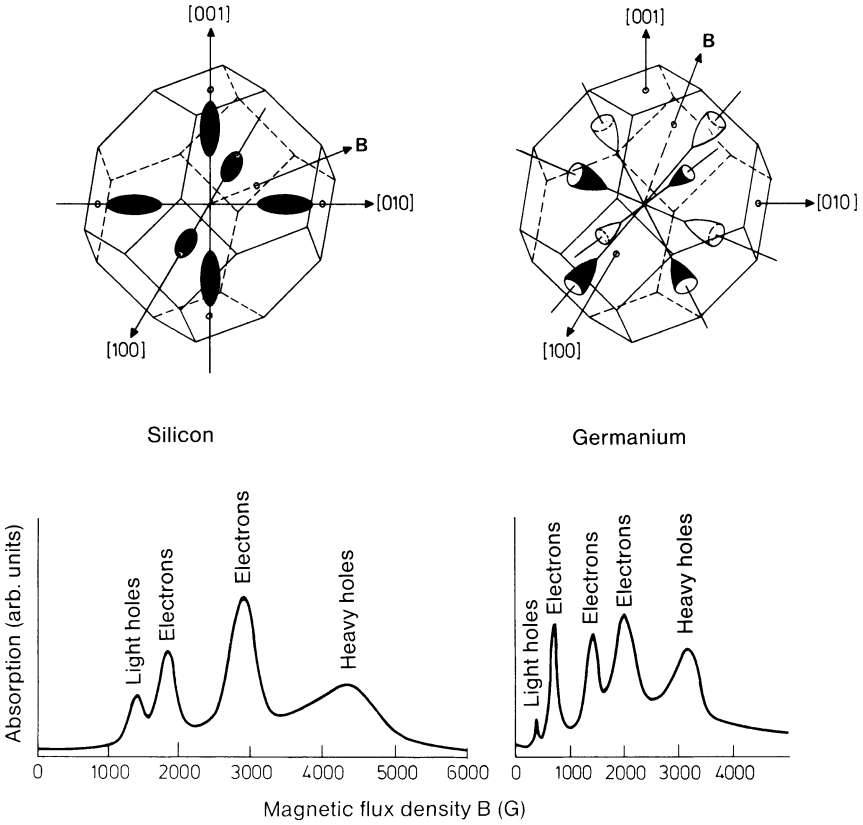

Fig. XV.2. (**a**) Surfaces of constant energy near the minimum of the conduction band for Si (*left*) and Ge (*right*). Electrons move on these surfaces along closed paths perpendicular to the magnetic field. The frequencies of the various orbits for a particular magnetic field are generally different. When the energy surfaces take the special form of ellipsoids, the frequencies are the same for all electrons on *one* ellipsoid. This, however, is unimportant for the observability of cyclotron resonance, since in each case it is the extremal orbit (here the orbit with the largest cross-sectional area) that leads to a clear maximum in the rf absorption. For an arbitrary orientation of the magnetic field there are three different extremal orbits for Si and four for Ge. (**b**) Cyclotron resonance absorption for Si (*left*) and Ge (*right*) with the magnetic field oriented in the (110) plane and angles of 30° and 60° respectively to the [100] direction (magnetic flux density measured in Gauss: $1\ \mathrm{G} = 10^{-4}\ \mathrm{T} = 10^{-4}\ \mathrm{Vs/m^2}$). In both cases, two of the extremal paths are equal for symmetry reasons, such that only two (three) remain distinct. The two valence bands at Γ with their different curvatures show up in the form of the peaks labeled "light" and "heavy" holes. For the observation of cyclotron resonance low temperatures are necessary. The electron-hole pairs were created by irradiation with light. (After [XV.1])

tions, two of the ellipsoid pairs have the same orientation with respect to the magnetic field and the number of absorption maxima for electrons is reduced from three (four) to two (three) for Si (Ge).

Cyclotron resonance is also observed for holes. The direction of their orbit, however, is opposite to that of electrons. It is thus possible to distinguish between electrons and holes by using a circularly polarized high-frequency field incident along the axis of the $\boldsymbol{B}$ field. Electrons and holes then absorb, according to the direction of their orbit, only right or left circularly polarized radiation. In the absorption spectra (Fig. XV.2 b), it can be seen that two maxima are assigned to holes. The interpretation here is different from the case of electrons. Hole states are found at the valence band maximum at the Γ point of the Brillouin zone. In a (simple) treatment, one would expect only spherical energy surfaces on the grounds of the symmetry of the Brillouin zone, that is to say only one cyclotron frequency. As discussed in Sect. 12.1, however (see also Figs. 12.3, 12.4), the band structure at the valence band maximum is rather more complicated: there are both "light" and "heavy" holes.

Cyclotron resonance can only be observed if the carriers are able to complete a large number of closed orbits around the magnetic field prior to suffering collisions with phonons or impurities. This is possible when the relaxation time τ (of the order of the mean time between collisions) is large compared to the inverse cyclotron frequency ω_c^{-1}. Cyclotron resonance is therefore only observed in pure and highly perfect crytals at low (liquid-He) temperature.

Reference

XV.1 G. Dresselhaus, A. F. Kip, C. Kittel: Phys. Rev. **98**, 368 (1955)

Panel XVI
Shubnikov-de Haas Oscillations and Quantum Hall Effect

Precise semiconductor heterostructures and superlattices can be prepared by using atomically controlled epitaxy (Sect. 12.7). Such structures allow the confinement of free conduction electrons in quasi-two-dimensional (2D) quantum-well structures with typical vertical dimensions between 20 and 100 Å. Thereby, quasi-2D electron gases (2DEG) are generated that exhibit highly interesting physical effects, in particular in externally applied strong magnetic fields. An electron gas in a 2D quantum well is confined in one direction z, perpendicular to the hetero-interface or to the layer sequence in a superlattice. Electrons are free to move about only in the x,y-plane perpendicular to the z-direction. Accordingly, the energy eigen-values of an electron within the 2DEG are given as (Sect. 12.7):

$$E_j(\boldsymbol{k}_\parallel) = \frac{\hbar^2}{2m_x^*}k_x^2 + \frac{\hbar^2}{2m_y^*}k_y^2 + \varepsilon_j \,, \quad j = 1, 2, \ldots \tag{XVI.1}$$

where m_x^* and m_y^* are the effective mass components for a free motion in the x,y-plane. The discrete eigen-values that result from the quantization in the z-direction are denoted by ε_j (12.90). For a rectangular quantum well of thickness d_z (Sect. 6.1) with infinitely high barriers, e.g., ε_j correspond to standing electron waves in a potential box (12.93).

$$\varepsilon_j \simeq \frac{\hbar^2\pi^2}{2m_z^*}\,\frac{j^2}{d_z^2} \,, \quad j = 1, 2, 3, \ldots \,. \tag{XVI.2}$$

Equation (XVI.2) thus describes a set of discrete energy parabolas along k_x and k_y, the so-called subbands (Sect. 12.7). Application of a strong magnetic field $\boldsymbol{B}$ perpendicular to the x,y-plane of free-electron movement further decreases the dimensionality of the electron gas. The electrons are forced into cyclotron orbits perpendicular to the $\boldsymbol{B}$-field (Panel XV). The cyclotron frequency (frequency of circulation)

$$\omega_c = -\frac{eB}{m_\parallel^*} \tag{XVI.3}$$

is determined by the balance between the Lorentz force and the centrifugal force. $m_\parallel^*$ is the effective mass of the electrons parallel to the orbital movement (for isotropy in the x,y-plane: $m_x^* = m_y^* = m_\parallel^*$). Since an orbital movement can

be decomposed into two linear harmonic oscillations perpendicular to each other, the quantum-mechanical energy eigenvalues of the orbits are those of a harmonic oscillator with eigenfrequency ω_c. The magnetic field $\boldsymbol{B}$ therefore causes a further quantization of the subbands with single particle energies:

$$E_{j,n,s} = \varepsilon_j + (n + \tfrac{1}{2})\hbar\omega_c + sg\mu_B B\,. \qquad \text{(XVI.4)}$$

The last term accounts for the two spin orientations in the magnetic field in which $s = \pm 1$ is the spin quantum number, μ_B the Bohr magneton and g the Landé factor of the electron. The magnetic-field-induced quantization (XVI.4) has already been derived for the free electron gas of a metal by a somewhat different method in Panel VIII. In both cases the $\boldsymbol{B}$-field-induced quantization into so-called Landau levels (energetic separation $\hbar\omega_c$) leads to a splitting of the continuous energy parabolas (subbands) into discrete energy eigenvalues (Fig. XVI.1 b). Because of the 2D-dimensionality of the bands the density of states at vanishing magnetic field ($\boldsymbol{B} = \boldsymbol{0}$) is a step function (Fig. 12.28 c) for each particular subband. In a finite field $\boldsymbol{B} \neq \boldsymbol{0}$ this continuous function splits into a series of δ-like peaks that are separated

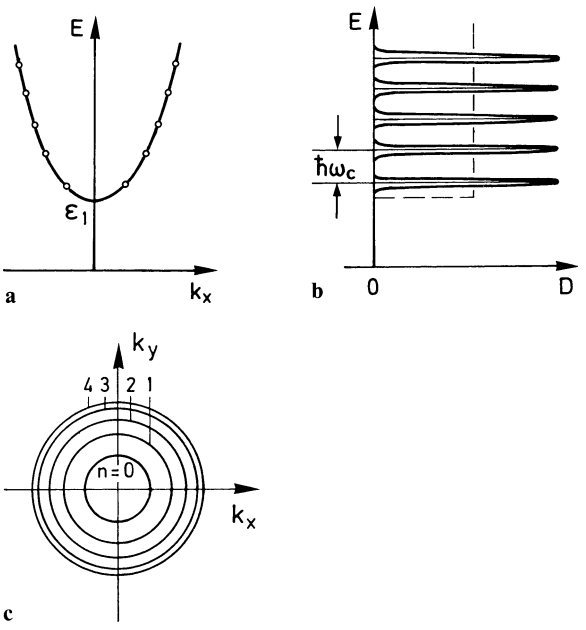

Fig. XVI.1 a–c. Qualitative illustration of the quantization of a 2D electron gas in an external magnetic field $\boldsymbol{B}$ (perpendicular to the (x, y) plane of free-electron propagation). (**a**) Energy parabola of the first subband (12.92) of a free 2D electron gas along k_x. A further quantization in the form of Landau states (*points*) appears upon application of an external magnetic field; (**b**) density of states D in the first subband of the 2D electron gas, without a magnetic field (*dashed line*) and with an external magnetic field (*solid lines*), $\omega_c = -eB/m_{\parallel}^{*}$ is the cyclotron frequency of the electron orbits perpendicular to the magnetic field $\boldsymbol{B}$; (**c**) representation of the Landau splitting $\left(n + \frac{1}{2}\right)\hbar\omega_c$ in the (k_x, k_y) plane of the reciprocal space

from each other by the energy $\hbar\omega_c$. The quantum states "condensate" into sharp Landau levels. Since the number of states is conserved in the condensation, a δ-like Landau level must contain exactly the number of states that were originally, at $\boldsymbol{B} = \boldsymbol{0}$, distributed over the energy range between two neighboring Landau levels, i.e. the degeneracy of a Landau level amounts to

$$N_L = \hbar\omega_c D_0 \,, \tag{XVI.5}$$

where D_0 is the state density of the subband at $\boldsymbol{B} = \boldsymbol{0}$. As opposed to (12.96) the spin degeneracy is lifted in the magnetic field and consequently the density of states is a factor of 2 smaller than in (12.96):

$$D_0 = m_{\parallel}^*/2\pi\hbar^2 \,. \tag{XVI.6}$$

Thus the degeneracy of a Landau level is

$$N_L = eB/h = 2.42 \times 10^{10}\,\mathrm{cm^{-2}T^{-1}} \cdot B \,. \tag{XVI.7}$$

A Landau level with an energy below the Fermi level is occupied with N_L electrons at sufficiently low temperature. A variation of the external magnetic field changes the energetic separation between the Landau levels, as well as their degree of degeneracy (XVI.5). With increasing magnetic field strength, the Landau levels shift to higher energies and finally cross the Fermi energy E_F; the Landau levels are emptied and the corresponding electrons occupy the next lower Landau level. This becomes possible because of the increased degree of degeneracy (XVI.7). In the case of a sharp Fermi edge at a sufficiently low temperature, the electron system reaches its lowest free energy each time when a Landau level has just crossed the Fermi level. With increasing $\boldsymbol{B}$-field the free energy increases again, until the next Landau level is emptied. As a consequence, the free energy oscillates as a function of the external magnetic field, and so do the material constants, such as the conductivity (Shubnikov-de Haas effect) and the magnetic susceptibility (de Haas-van Alphen effect). The de Haas-van Alphen effect has been of considerable importance in studies of the topology of Fermi surfaces in metals (Panel VIII).

The quantization of states in a magnetic field also gives rise to the quantum Hall effect in semiconductors with 2D-electron gases (2DEG). Since its discovery by Klaus von Klitzing (Nobel prize 1985) the quantum Hall effect [XVI.1] has become an important tool to characterize heterostructures in semiconductor physics. Moreover, the quantum Hall effect had an important impact on the development of the physics of nanostructures in general.

The experimental arrangement for the observation of the quantum Hall effect is similar to that of the classical Hall effect (Panel XIV): a Hall voltage U_H is measured perpendicular to the current through the sample at two opposite contacts. Unlike in the classical Hall-effect, however, the current is carried by a 2DEG in a semiconductor heterostructure, e.g., in a modulation-doped AlGaAs/GaAs heterostructure. The magnetic field $\boldsymbol{B}$ is oriented normal to the current flow and also normal to the plane of the 2DEG

(insert in Fig. XVI.2 and Fig. XVI.5c). In order to ensure a sharp Fermi edge in the highly degenerate 2DEG the measurement is performed at low temperature. When the magnetic field $\boldsymbol{B}$ is increased the Hall resistance $r_H = U_H/I$ varies stepwise, with steps at those B values where a sharp Landau level crosses the Fermi energy (Fig. XVI.2a). A simultaneous measurement of the magnetoresistance ϱ_{xx} ($\propto U_L/I$) parallel to the current flow via a pair of contacts placed along the direction of the current yields a sequence of sharp resistance peaks each time when in the Quantum Hall effect r_H jumps from one plateau to the next, i.e. where a Landau level crosses the Fermi energy E_F. For B values in between, ϱ_{xx} is negligible, at least for higher magnetic fields. These resistance oscillations are the Shubnikov-de Haas oscillations of a 2DEG.

The experimentally observed values of the Hall resistance at those magnetic fields where Landau levels cross E_F are derived directly from the relation for the classical Hall effect (Panel XIV). From (XIV.3) one obtains

$$r_H = \frac{U_H}{I} = \frac{-B}{nde} = \frac{-B}{eN_{2D}} \,, \qquad \text{(XVI.8 a)}$$

where N_{2D} is the two-dimensional carrier density of the 2DEG which, of course, amounts to multiples of the degree of degeneracy N_L (XVI.7) in a magnetic field $\boldsymbol{B}$.

The Hall resistance at Landau level crossing E_F are thus obtained as:

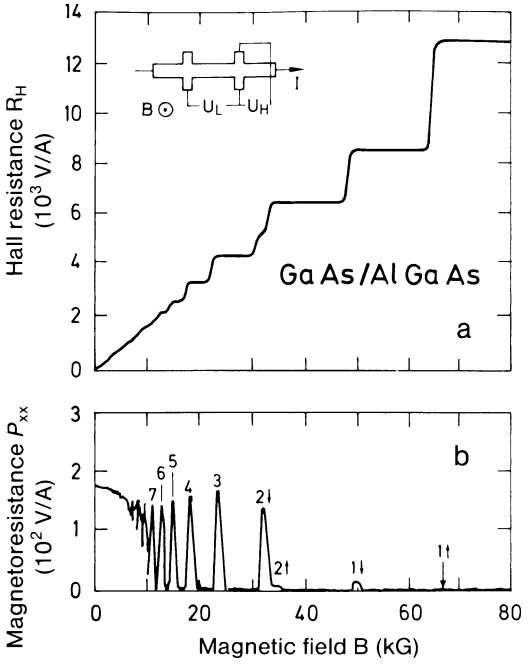

Fig. XVI.2. (**a**) Quantum Hall effect measured at 4 K for the quasi-2D electron gas of a modulation-doped AlGaAs/GaAs heterostructure; 2D electron density $N = 4 \times 10^{11}$ cm^{-2}, electronic mobility $\mu = 8.6 \times 10^4$ cm^2/Vs. The Hall resistance $R_H = U_H/I$ was measured as a function of the magnetic field $\boldsymbol{B}$ using the arrangement shown in the inset; (**b**) Shubnikov-de Haas oscillations of the magnetoresistance ϱ_{xx}, measured as indicated in the inset via U_L/I as a function of the external magnetic field $\boldsymbol{B}$. The numbers denote the subbands and the arrows indicate the spin direction relative to the $\boldsymbol{B}$ field (after [XVI.2])

$$r_H = \frac{-B}{\nu N_L e} = -\frac{h}{e^2}\frac{1}{\nu}\,, \quad \nu = 1, 2, 3, \ldots \qquad \text{(XVI.8 b)}$$

(XVI.8 b) describes correctly single discrete points of the experimental curve in Fig. XVI.2 a, but the characteristic plateaus in the Hall resistance are not at all explained. In order to understand the causes for the existence of plateaus a more profound consideration of the transport in a DEG in the presence of magnetic fields is necessary. In a strong magnetic field, electrons in the 2DEG are forced into cyclotron orbits. Undisturbed, closed orbits, however, are possible only in the interior of the sample (Fig. XVI.3). In that

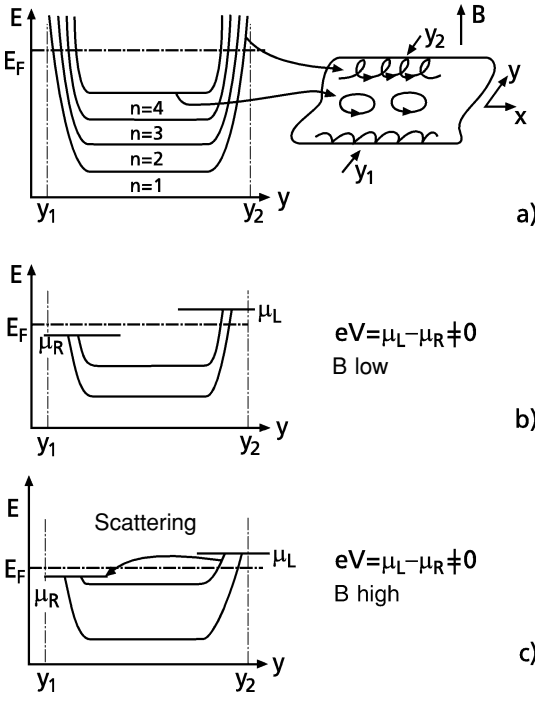

Fig. XVI.3. Explanation of the so-called edge channels for electrons when a magnetic field $\boldsymbol{B}$ is applied perpendicular to the plane of a 2D electron gas (2DEG). (**a**) A strong magnetic field $\boldsymbol{B}$ perpendicular to the 2DEG causes quantization of the electronic states into Landau levels ($n = 1, 2, 3, \ldots$), which correspond to closed cyclotron orbits in the interior of the sample. At the sample boundaries y_1 and y_2 orbiting is interrupted by elastic reflection from the surface. The increased spatial confinement entails a strong enhancement of the energy of the Landau levels, which eventually cross the Fermi energy E_F. (**b**) Under the condition of ballistic current flow the chemical potentials μ_R and μ_L of the right and left edge channels are different and determine the potentials of the corresponding right and left contacts. (**c**) With increasing magnetic field $\boldsymbol{B}$ the splitting between the Landau levels increases, while the upper Landau level approaches the Fermi level E_F. Electronic states near E_F in the interior of the sample become available for scattering processes and the left and right edge channels communicate by these scattering processes

case, the orbital energies are those of the undisturbed Landau levels (XVI.4). Electrons near the boundaries of the sample on the other hand are hindered to complete a full cyclotron orbit. In the simplest case of an elastic mirror reflection at the boundary the electron path becomes a sequence of partial circles (skipping orbits). The combination of cyclotron orbits and surface scattering in the skipping orbits amounts to an additional spatial confinement of the electrons (Fig. XVI.4). By virtue of the uncertainty principle, a stronger confinement is equivalent to a higher kinetic energy. Near the boundaries of the sample the energies of the Landau level are therefore bent upwards and eventually even cross the Fermi level E_F at y_1 and y_2 near the sample surface (Fig. XVI.3 a). Crossing the Fermi level means that one has metallic conduction. Thus, a Landau level that cannot contribute to the electrical conductance in the bulk of the material (since its energy is deep below the Fermi level) does contribute to the conductance at the surface, in the so-called edge channels. The edge channels possess a further important property. Transport in these states is quasi-ballistic, even in macroscopic samples containing impurities. When a carrier in an edge channel is scattered from a defect S (Fig. XVI.4) its trajectory is redirected into the forward direction by the Lorentz force of the strong magnetic field. The total current in the forward direction is therefore not reduced by the presence of impurities and transport in edge channels is therefore quasi-ballistic, i.e. without resistance, regardless of the presence of impurities.

The two edge channels on either side are electrically isolated from each other provided that the Landau levels in the interior of the sample are more than a few $k_B T$ below the Fermi energy E_F. Because of the quasi-ballistic transport in the channels, electrons within one channel are at the same potential. For an electron current flow from left to right this potential is the potential μ_L of the left contact, while electrons in the right "backwards" edge channels are at the potential of the right contact μ_R (Fig. XVI.5). The difference in the potential is the applied voltage multiplied by the electron charge

$$\mu_L - \mu_R = eV\,. \tag{XVI.9}$$

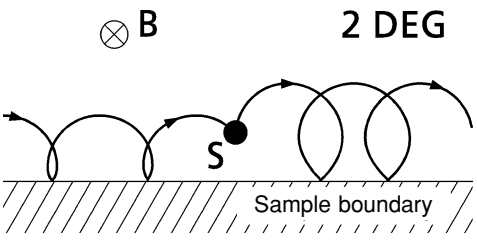

Fig. XVI.4. Schematic trajectory of a carrier in an edge channel that undergoes an elastic scattering process on a defect atom S. The magnetic field is oriented perpendicular to the 2DEG. The 2DEG is assumed to be confined by the sample boundary. The scattering processes at the boundary are assumed to be elastic

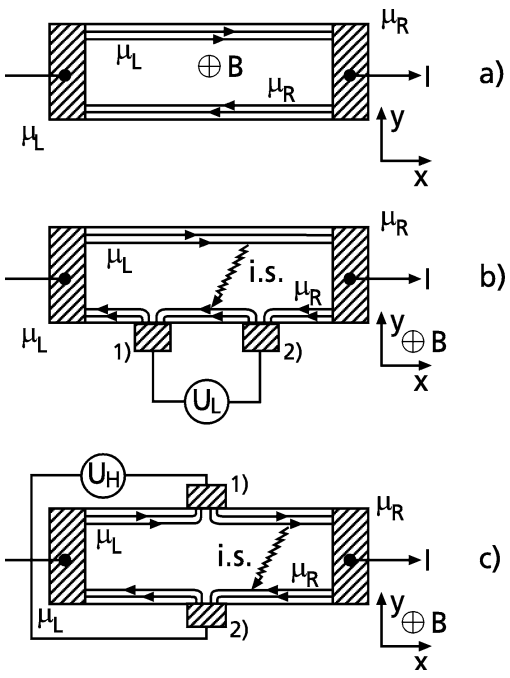

Fig. XVI.5. (**a**) Ballistic electron transport in edge channels that are generated by quantization in a spatially limited 2 DEG due to a strong magnetic field ***B*** perpendicular to the 2DEG. Under the assumption of negligible inelastic scattering between left (L) and right (R) channels the chemical potentials μ_L and μ_R of the left and right channels equal those of the left and right electrical contacts. (**b**) Conduction through edge channels in a 2 DEG in the presence of two electrical contacts 1) and 2) for measuring the Shubnikov-de Haas oscillations in a strong magnetic field. Inelastic scattering (i.s.) between left and right channels causes a potential difference between the contacts. (**c**) Conduction through edge channels in a 2 DEG with two opposite contacts 1) and 2) for measuring the quantum Hall effect

The net electric current carried by the left and right (1D) edge channels is the difference between the contributions from both contacts (see also Sect. 9.9, (9.96)):

$$I = \sum_{n=1}^{n_c} \int_{\mu_R}^{\mu_L} e D_n^{(1)}(E) v_n(E) dE \,. \tag{XVI.10}$$

The sum runs over all occupied Landau levels n (up to a maximum occupied level n_c), which cross E_F as edge channels, $D_n^{(1)} = (2\pi)^{-1}(dE_n/dk_x)^{-1}$ is the 1D density of states of the n-th edge channel, while $v_n = \hbar^{-1}(dE_n/dk_x)$ is the electron velocity within this channel. The n-th edge channel thus contributes to the current an amount

$$I_n = \frac{e}{h}(\mu_L - \mu_R) = \frac{e^2}{h} V \,. \tag{XVI.11}$$

The quantum Hall effect is measured as the resistance between left and right edge channels via two contacts perpendicular to the current flow (Fig. XVI.5c). According to (XVI.11) each channel contributes an amount of e^2/h to the total conductance, i.e. h/e^2 to the Hall resistance. This is exactly

the quantum (XVI.8 b) by which the resistance is stepwise enhanced each time a further Landau level crosses the Fermi energy and is thus depleted of electrons. The Shubnikov-de Haas effect is explained in a similar way. In this case, the magneto-resistance is measured along the path of the current via two contacts arranged along the edge channels (Fig. XVI.5 b). As long as the Landau levels in the interior of the sample are sufficiently far away from E_F, no voltage drop occurs, because of the ballistic transport. At a particular magnetic field one of the Landau levels in the interior of the sample reaches the Fermi level E_F and electronic states for scattering between the left and right edge channel become available. Scattering between the forth and back channels induces a resistance along the channel direction. Hence, at these particular magnetic fields a peak-like increase of the resistance is observed (Fig. XVI.2 b). In this interpretation, both quantum Hall effect and Shubnikov-de Haas oscillations are attributed to quasi-one-dimensional carrier transport in edge channels [XVI.4]. The step-like change of the Hall resistance r_H can be determined experimentally with an accuracy of 10^{-7} on samples of different structure and from different groups [XVI.1]. The measurement of quantum Hall effect is therefore presently used for a very precise determination of the fine structure constant.

$$\alpha = \frac{e^2}{h}\frac{\mu_0 c}{2} \cong \frac{1}{137} \ . \tag{XVI.12}$$

If, on the other hand, α is assumed to be known (α is determined by a large number of independent precise experiments) the quantum Hall effect allows the introduction of a standard for the resistance unit "Ohm".

It should finally be mentioned that the quantum Hall effect has the same physical origin as a number of other phenomena related to 1D transport in mesoscopic systems and nanostructures. The fact that edge channels with their 1D conductance (induced by the high magnetic field) each contribute an amount of e^2/h to the total conductivity can directly be transferred to other types of 1D conductance channels. The only requirement is that the width of the channels is comparable with the Fermi wavelength λ_F of the electrons and that the transport is quasi-ballistic, i.e. that the sample dimensions (wire length) are smaller than the mean free path of the carriers. It was, for example, shown that short 1D channels (induced by two metal gates evaporated on the sample) within a 2D electron gas at a modulation-doped AlGaAs/GaAs heterostructure (Sect. 12.7) show a step-like variation of their electrical conductance with varying gate voltage. The channel conductance increases by quanta of $2e^2/h$ when the channel width is increased by a positive bias on the metal gate contacts on top [XVI.5]. The conductance quantum amounts to $2e^2/h$ because of spin degeneracy (see Sect. 9.9).

References

XVI.1 K. von Klitzing, G. Dorda, M. Pepper: Phys. Rev. B **28**, 4886 (1983)
XVI.2 M.A. Paalonen, D.C. Tsui, A.C. Gossard: Phys. Rev. B **25**, 5566 (1982)
XVI.3 H. Lüth: *Solid Surfaces, Interfaces and Thin Films* (Springer, Berlin Heidelberg 2001) 4th edition, p. 419
XVI.4 M. Janßen, O. Viehweger, U. Fastenrath, J. Hajdu: *Introduction to the Theory of the Integer Quantum Hall Effect* (VCH, Weinheim 1994)
XVI.5 B.J. van Wees, H. van Houten, C.W.J. Beenakker, J.W. Williamson, L.P. Kouwenhoven, D. van der Marel, C.T. Foxon: Phys. Rev. Let. **60**, 848 (1988)

Panel XVII
Semiconductor Epitaxy

In modern semiconductor physics and device technology, thin crystalline layers are playing an ever increasing role. Of particular interest are multilayer structures, for example, multiple layers of AlGaAs and GaAs, or layers of GaInAs on an InP substrate. Such systems enable one to study novel phenomena such as the quantum Hall effect (Panel XVI), and, at the same time, are the basis for the production of fast transistors and semiconductor lasers. Semiconductor layer structures are almost invariably produced by means of epitaxy. This is a method that nowadays allows extremely precise control of the growth – so precise that single layers of atoms can be reliably deposited under favorable conditions.

Samples for use in fundamental research are most commonly produced by ***M**olecular **B**eam **E**pitaxy* (MBE), a process in which semiconductor layers are deposited epitaxially on a substrate in an ultra-high-vacuum chamber [XVII.1, 2] (Fig. XVII.1). The principle behind the method is simple: a substance such as Ga or Al is evaporated and the vapor deposited on a substrate such as GaAs. Ultra-high-vacuum (UHV) systems with base pressures in the 10^{-8} Pa range are used to prevent contamination and to ensure well-defined conditions, both in the molecular beam and at the surface of the substrate. One can estimate that, at a pressure of 10^{-8} Pa, it will take a few hours for a newly prepared surface to become covered with a monolayer of adsorbates, even if every impinging gas molecule sticks to the surface. Such UHV chambers are made of stainless steel, and the extremely low pressures of less than 10^{-8} Pa (corresponding to a molecular mean free path of some meters) are maintained by ion-getter pumps or turbomolecular pumps. In an ion-getter pump, gas atoms are ionized by high electric fields and then adsorbed onto appropriately charged active metal films (e.g. Ti) – this latter process being known as "gettering". The pumping action of a turbomolecular pump relies on the momentum exchange between gas molecules and the blades of a rapidly rotating turbine.

The UHV chamber that comprises the main part of an MBE machine (Fig. XVII.1) is equipped with an internal cryoshield cooled by liquid nitrogen. This serves to trap stray atoms and molecules, thereby further reducing the background pressure. The substrate material, e.g. a GaAs wafer, onto which an AlGaAs layer is to be deposited, is mounted on a rotating substrate holder. During growth the substrate must be maintained at a temperature of about 500–600 °C in order to ensure a sufficiently high mobility of the imping-

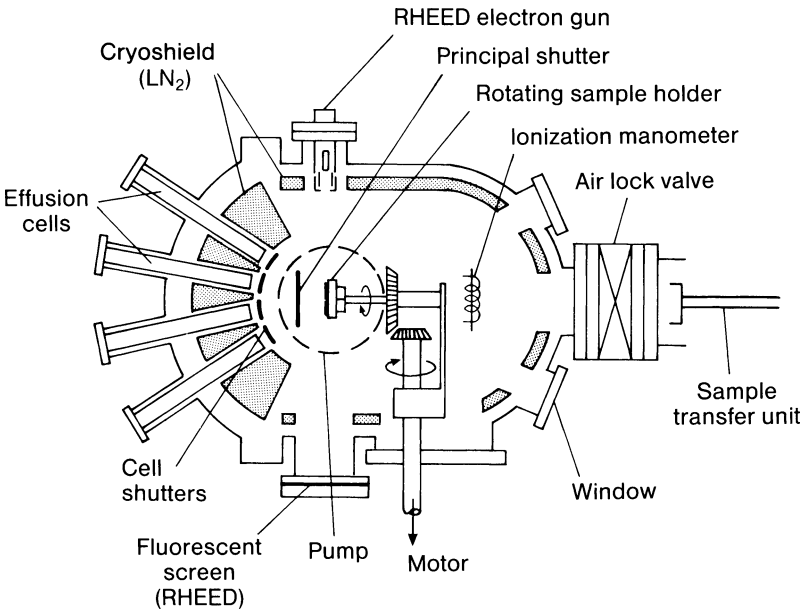

Fig. XVII.1. Schematic illustration of a UHV chamber for molecular beam epitaxy (MBE) of III–V semiconductor layers (viewed in section from above). (After [XVII.3])

ing atoms or molecules (Ga, As_4, Si, etc.) on the surface. The atoms or molecules to be incorporated into the growing crystal originate from a so-called effusion source in which the starting material, e.g. solid Ga and As for GaAs epitaxy, is vaporized from boron-nitride crucibles by electrical heating. Mechanically driven shutters, controlled from outside, can open and close the individual effusion cells, thereby switching the corresponding molecular beams on and off. The growth rate of the epitaxial layer is determined by the particle flux in the molecular beam, which, in turn, is controlled by varying the temperature of the crucible. To obtain a particular sequence of well-defined layers with atomically sharp boundaries, as is required for a composition superlattice (Sect. 12.7), the crucible temperature and the shutter-opening times must be regulated by a computer program. This is particularly necessary for the epitaxial growth of ternary or quaternary alloys such as $Al_{0.55}Ga_{0.45}As$ with a fixed composition. Homoepitaxy of GaAs layers on a GaAs substrate does not demand any precise regulation of the molecular flux. In this case, nature herself provides a helping hand: stoichiometric growth of GaAs is possible even for a non-stoichiometric beam composition. The growth rate is determined solely by the rate of arrival of Ga atoms. At a substrate temperature of between 500 and 600°C, arsenic atoms only stick to the surface when sufficient numbers of Ga atoms are present on the growing surface; the sticking coefficient for As approaches one for a Ga excess on the surface and is close to zero when Ga is underrepresented. Thus GaAs epitaxy occurs optimally for an excess of arsenic. For a deeper understanding of the GaAs growth in MBE, it is also

important to know that the arsenic molecular beam produced by evaporation of solid arsenic consists mostly of As_4 molecules, which must be dissociated into As atoms by the thermal energy of the hot substrate surface before they can be incorporated in the growing layer. This dissociation process is sometimes carried out thermally in hot graphite "cracker" cells, which leads to better layer quality. Important dopants in III–V MBE are Si for *n* and Be for *p* doping. These materials are also evaporated from effusion cells which can be switched off and on in a controlled way by the opening and closing of shutters.

As shown in Fig. XVII.1, an MBE system is normally also furnished with an electron gun and a fluorescent screen for the observation of diffraction patterns (***R**eflection **H**igh **E**nergy **E**lectron **D**iffraction*, RHEED). Diffraction patterns provide information about the crystallographic structure of the growing surface. An ionization pressure gauge is also included to measure both the chamber pressure and the pressure in the molecular beams.

The MBE system shown in Fig. XVII.1 is of the type frequently employed for the epitaxy of III–V and II–VI semiconductors. Si MBE [XVII.4] is carried out in similar systems, but the Si molecular beam is not produced by thermal evaporation from effusion cells, but rather by electron bombardment of a solid Si target, which thereby becomes so hot that it evaporates.

The decisive advantage of MBE over other epitaxial methods which work at higher presssure is the fast switching time between different sources. This provides a straightforward method of obtaining sharp doping and composition profiles. A typical growth rate is 1 μm/h, which corresponds to about 0.3 nm/s, or one monolayer per second. The time needed to switch between different sources should therefore be well below one second.

Such short switching times are more difficult to achieve using the second important epitaxial method, ***M**etal **O**rganic **C**hemical **V**apor **D**eposition* (MOCVD) [XVII.5]. Here growth takes place in a reactor, whose whole gas volume must be exchanged on switching from one source to another. Compared to MBE, MOCVD has the interesting advantage for industrial applications that the sources can be easily controlled by gas flow regulators. Furthermore the gas phase sources allow almost continuous operation of the system.

We will discuss the principles of MOCVD processes using GaAs epitaxy as an example. We are concerned here with the deposition of solid GaAs from gas phase materials containing Ga and As. The molecule AsH_3 and the metal organic gas trimethylgallium [$TMG = Ga(CH_3)_3$] are frequently used. The overall reaction, which takes place via several complicated intermediate steps, can be written

$$[Ga(CH_3)_3]_{gas} + [AsH_3]_{gas} \rightarrow [GaAs]_{solid} + [3CH_4]_{gas} \; . \qquad \text{(XVII.1)}$$

AsH_3 is fed directly from a gas bottle via a gas regulating valve into the quartz reactor (Fig. XVII.2a). The metal organic component TMG is in a bulb, and its vapor pressure is controlled by a temperature bath. Hydrogen (H_2) flows through this bulb and transports the TMG to the reactor. Another gas line allows the whole system to be flushed with H_2. The

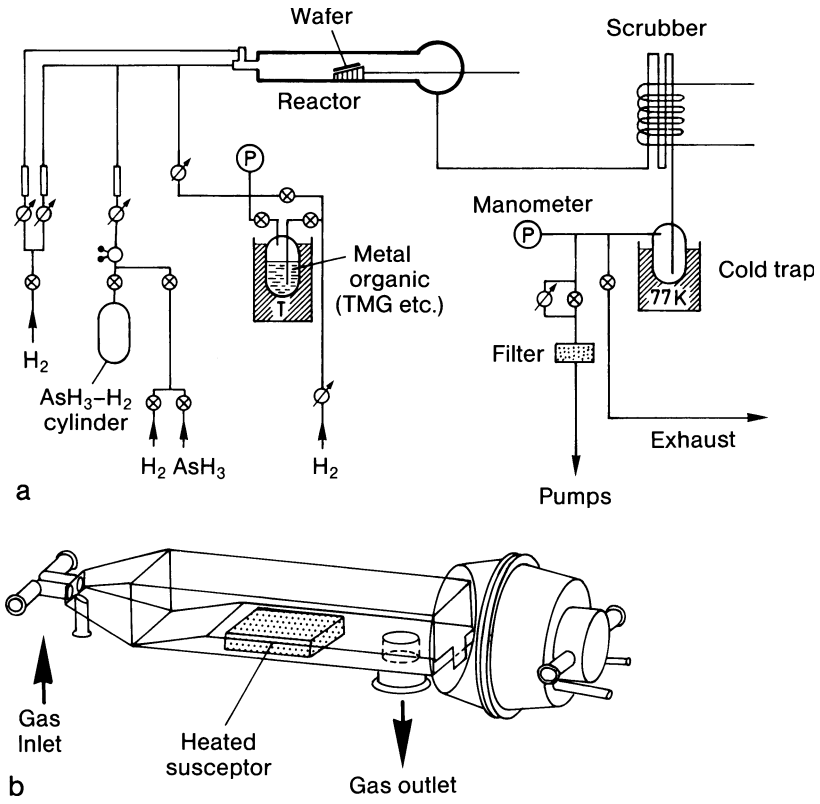

Fig. XVII.2 a, b. Schematic illustration of an apparatus for metal organic chemical vapor deposition (MOCVD) of III–V semiconductor layers. (**a**) General overview; (**b**) quartz glass reactor (typical length 50 cm); the susceptor, which holds the wafer to be epitaxially coated, is heated during growth to 700–1200 K. (After [XVII.6])

components which are not consumed in the reaction, and the reaction products, are pumped out at the end of the reactor through a decomposition oven, to eliminate the dangerous excess AsH_3. The pumping system also allows the reactor to be operated at low pressure (so-called low pressure MOCVD). With this method the switching time between different sources can be reduced and most of the advantages of MBE obtained. Besides the sources for the growth of the base material (AsH_3 and TMG) other gas lines are needed for the doping gases, such as SiH_4, $(C_2H_5)_2Te$, $(C_2H_5)_2Mg$ for Si, Te or Mg doping. For the growth of ternary and quaternary III–V alloys further gas lines are also necessary. These supply metal organic materials such as trimethylaluminium $(CH_3)_3Al$, trimethylantimony $(CH_3)_3Sb$, trimethyl-indium $(CH_3)_3In$ and others. For the growth of phosphorus-containing materials such as InP and GaP, the hydride phosphine PH_3 is used. Figure XVII.2 b shows a possible design for a flux reactor. The wafer on which the films are to be deposited, the so-called graphite susceptor, is

held at a temperature between 600 and 800°C during growth. Heating takes place by radiation, direct electrical current or microwave heating.

Compared with MBE, the growth process in MOCVD is considerably more complicated. A stream of gas flows over the growing layer, and from this stream the reaction components diffuse to the surface. Both at the surface and in the gas phase dissociation reactions take place, e.g. AsH_3 is dissociated by collision in the gas phase as well as on the surface itself. After the required surface reactions have occurred, e.g. the incorporation of the dissociated As into the growing crystal lattice, the reaction products such as CH_4 are again transported away from the surface by the gas stream. MOCVD growth is therefore largely determined by transport to and from the surface and by surface reaction kinetics. This is seen clearly if one plots the rate of growth in the MOCVD process as a function of the substrate temperature (Fig. XVII.3). The result is a typical curve in which the fall-off at lower temperatures has the form $\exp(-E_a/kT)$, since the rate is kinetically limited by surface reactions. Typical activation energies, E_a, in this region are around 1 eV per atom. The temperature range in which this kinetically limited region occurs depends on the source material. For the relatively stable TMG, the exponential decrease is observed at temperatures below 850 K, but for the more easily dissociated triethyl-gallium [TEG, $(C_2H_5)_3Ga$] it occurs below 700 K. For temperatures above the kinetically limited region, the growth rate shows a plateau whose height depends on the conditions influencing diffusion to and from the substrate (e.g. the flow velocity in the reactor). At still higher temperatures, another decrease in the growth rate is observed (Fig. XVII.3). It is probable that the reasons for this

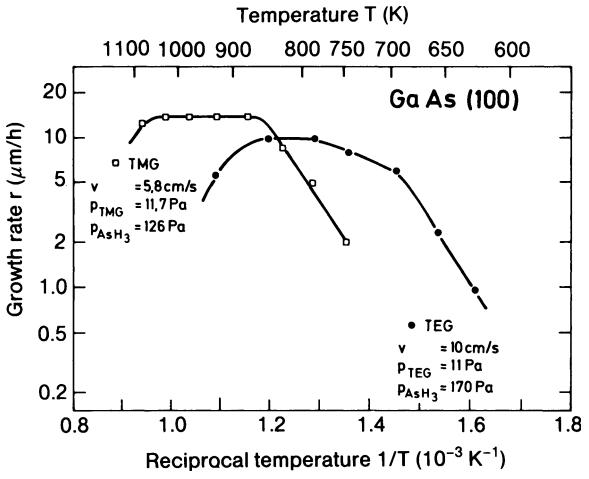

Fig. XVII.3. Growth rates of GaAs layers during metal organic chemical vapor deposition (MOCVD) from the starting materials AsH_3 and trimethyl-gallium (TMG) or triethyl-gallium (TEG); GaAs wafer orientation (100). The gas flow velocities v and the partial pressures P_{TEG}, P_{TMG}, P_{AsH_3} are given. The total pressure in each experiment is 10^4 Pa. (After [XVII.7])

are not inherent to the process. A likely cause is the loss of reactants from the gas stream due to deposition on the reactor walls.

By clever process control in low pressure MOCVD, and by using specially designed valves, very short switching times can now be achieved. Thus, even with MOCVD it is possible to produce atomically sharp heterojunctions between two semiconductors.

A third modern epitaxial method, so-called ***M**etal **O**rganic* MBE (MOMBE), also called CBE (***C**hemical **B**eam **E**pitaxy*), combines the advantages of MBE and MOCVD [XVII.8]. As in MBE, a UHV system serves as the growth chamber (Fig. XVII.1). The source materials, however, are gases, as in MOCVD, and these are fed through lines with controlled valves into the UHV chamber. Inside the chamber, specially constructed admission systems (capillaries) serve to form the molecular beams, which are directed at the surface of the substrate to be coated. For the growth of GaAs, one might use, for example, AsH_3 and triethylgallium [$(C_2H_5)_3Ga$, TEG]. In a MOCVD reactor, collisions in the gas above the hot substrate surface lead to a significant predissociation of the AsH_3, but this does not occur here because of the very low background pressure. In MOMBE, the AsH_3 must therefore be thermally dissociated in the inlet capillary.

All metal organic epitaxy processes suffer from the presence of carbon impurities, which originate in the decomposition of the metal organics. The carbon is most frequently incorporated at As lattice positions and leads to *p* conduction (although it can also be incorporated as a donor at Ga sites). Carbon inclusion becomes stronger the lower the pressure used for epitaxy. Thus in MOMBE, GaAs layers with low levels of *p* doping can only be achieved using TEG, whereas in MOCVD it is also possible to use TMG.

References

XVII.1 M.A. Herman, H. Sitter: *Molecular Beam Epitaxy*, 2nd ed., Springer Ser. Mater. Sci., Vol. 7 (Springer, Berlin, Heidelberg 1996)

XVII.2 E.H.C. Parker (Ed.): *The Technology and Physics of Molecular Beam Epitaxy* (Plenum, New York 1985)

XVII.3 A.Y. Cho, K.Y. Cheng: Appl. Phys. Lett. **38**, 360 (1981)

XVII.4 E. Kasper, H.-J. Herzog, H. Dämbkes, Th. Richter: Growth Mode and Interface Structure of MBE Grown SiGe Structures, in *Two-Dimensional Systems: Physics and New Devices*, ed. by G. Bauer, F. Kuchar, H. Heinrich, Springer Ser. Solid-State Sci., Vol. 67 (Springer, Berlin, Heidelberg 1986)

XVII.5 Proc. ICMOVPE I, J. Crystal Growth **55** (1981); Proc. ICMOVPE II, J. Crystal Growth **68** (1984)
W. Richter: Physics of Metal Organic Chemical Vapour Deposition: *Festkörperprobleme* **26**, 335 (Advances in Solid State Physics 26) ed. by P. Grosse (Vieweg, Braunschweig 1986)

XVII.6 H. Heinecke, E. Veuhoff, N. Pütz, M. Heyen, P. Balk: J. Electron. Mater. **13**, 815 (1984)

XVII.7 C. Plass, H. Heinecke, O. Kayser, H. Lüth, P. Balk: J. Crystal Growth **88**, 455 (1988)

XVII.8 H. Lüth: in Surf. Science **299/300**, 867 (1994)

Panel XVIII
Preparation of Nanostructures

Studies on electronic transport in nanostructures require samples that are appropriately downscaled in their dimensions. For quantum transport as described in Sect. 9.9, e.g., cross sections in the range between 10 and 100nm are necessary. Suitable structuring techniques, for metals as well as for semiconductors, have been developed in the context of microelectronics, where they are used for the fabrication of highly integrated circuits on silicon wafers. One common method for the preparation of 3-dimensional nanostructures on a substrate begins with the deposition of the desired material on a large area substrate wafer, e.g. silicon, GaAs, or sapphire (Al_2O_3). Deposition of metals is frequently performed by evaporation from electrically heated tungsten or graphite crucibles or by sputter-deposition. Crystalline growth of the deposited epilayer is possible if the growing film is lattice-matched to the substrate. For ordered crystal growth, the substrate is typically heated in order to afford sufficient surface mobility to the deposited atoms. Growing crystalline epilayers is called epitaxy.

Epitaxy techniques such as MBE (**M**olecular **B**eam **E**pitaxy) or MOVPE/MOCVD (**M**etal**O**rganic **V**apour **P**hase **E**pitaxy/**M**etal**O**rganic **C**hemical **V**apour **D**eposition), (Panel XVII) are employed in particular for growing crystalline semiconductor layers and multilayer systems, e.g. of GaAs/AlGaAs, InGaAs/InP or Si/Ge on convenient substrates. Deposition of polycrystalline layers, amorphous layers or crystalline epilayers and multilayer systems thus enables the nanostructuring in one, namely the direction perpendicular to the substrate surface. For film growth rates of 1mm/h (i.e. 0.3nm/s), layer thicknesses below 1nm and atomically sharp interfaces are easily achieved.

In order to machine 3D nanostructures an additional lateral structuring of the layer systems is necessary. This is achieved by *lithography* (greek: $\lambda\iota\theta o\varsigma$ = stone, $\gamma\rho\alpha\varphi\epsilon\iota\nu$ = write). Lateral structuring down into the 100nm range is performed via *optical lithography*, i.e. by the use of illumination techniques with visible and UV light. Nanostructuring down to the 5-10nm scale presently requires *electron beam (e-beam) lithography* in electron optical columns. Optical and e-beam lithography are similar insofar as a *photoresist* is irradiated. The irradiation changes the chemical and structural properties of the resist such that its solubility in an organic solvent is modified. In optical lithography the irradiation of the resist is performed through a patterned mask, whether by contact illumination or by optical projection.

The desired pattern is thus transferred as a whole to the resist in parallel, i.e. in a single illumination step. Electron beam lithography, on the other hand, is a serial illumination technique, where a focused electron beam is scanned computer controlled over the resist film whereby the beam writes the pattern into the resist step-by-step. For this purpose, a high-precision electron microscope column is used, similarly as in a scanning electron microscope. In comparison with optical lithography, electron beam lithography is much more costly and time consuming. For industrial applications, present research aims therefore at replacing electron beam lithography by parallel illumination techniques with light of extremely short wavelength. An example is the so-called extreme UV (EUV) lithography, which uses coherent radiation pulses of *soft X-ray lasers*. The current most advanced laser of that type is based on the amplification of the spontaneous emission in a plasma of Ag-ions [XVIII.1]. A spatially coherent beam is generated by amplification in a second Ag-plasma. The output emission line with a wavelength of 13.9nm ($h\nu$ = 89.2eV) corresponds to the 4*d*-4*p* transition of nickel Ni-like Ag ions.

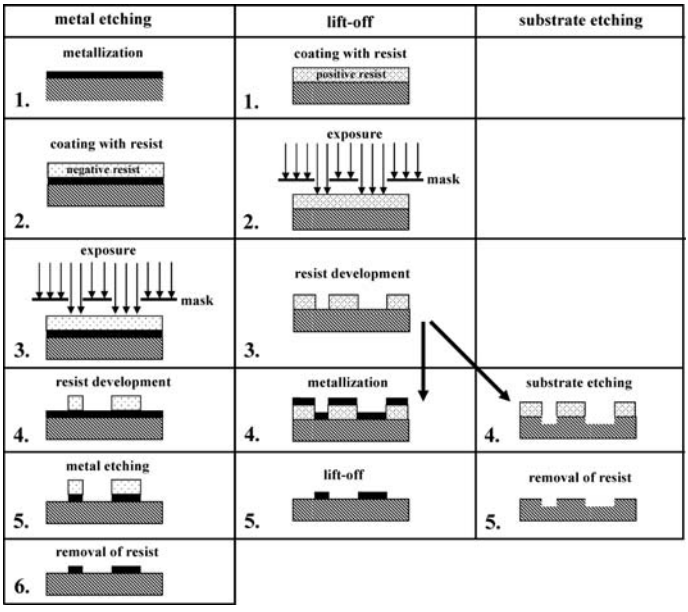

Fig. XVIII.1. Schematic presentation of important lithographical structuring techniques. The illumination steps for metal etching (3) and for the lift-off process (2) are performed in parallel in optical lithography by exposure through masks (as shown) or by projection, or alternatively in series by scanning electron beam lithography. A negative photoresist becomes insoluble in the illuminated areas while a positive resist, there, is soluble in the subsequent development step

In order to transfer lateral structures of the irradiated photoresist into semiconductor layer systems or onto a wafer, several etching processes are applied. The most important transfer methods are compiled in Fig. XVIII.1. In case of metal structuring on a wafer two techniques are commonly used, *direct metal etching* and the so-called *lift-off process*. In metal etching (left column in Fig. XVIII.1), patterns are etched into the metal film which has been deposited in advance by evaporation or sputtering (step 1). The metal film with a typical thickness of 100 nm is then spin-coated in a centrifuge with an organic photoresist (often PMMA = poly-methyl metacrylate) (step 2). In direct metal etching, a so-called negative resist is used, in which the locally illuminated areas – be it by means of light through a mask or by an electron beam (step 3) – exhibit a decreased solubility in the subsequent development process (step 4). The decreased solubility of the resist in the illuminated areas arises from irradiation induced polymer formation. In the development process (step 4) an organic solvent removes the non-illuminated areas of the resist layer. The remaining resist structures then protect the underlying metal film against metal etching (step 5). Depending on the requirements concerning edge sharpness the etching itself is performed by wet chemistry in solution (HCl, H_2O_2 etc.) or by ion bombardment in a plasma discharge (O_2), sometimes in combination with more or less chemically reactive species (HF, HBr). The latter technique, called ***R**eactive **I**on **E**tching* (*RIE*), is performed in vacuum chambers (RIE chamber, Fig. XVIII.2). After the metal etching, nanostructures consisting of metal and the resist on top remain on the substrate. In the last final step 6, the resist caps are removed by heating in an oven and the desired metal nanostructures are found on the wafer.

Since III-V semiconductors and metals exhibit similar etching properties, the hitherto described method is not well suited for the generation of metal-

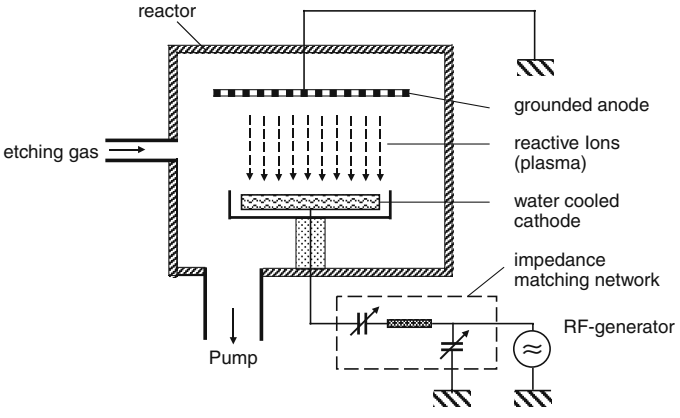

Fig. XVIII.2. Scheme of a set-up for Reactive Ion Etching (RIE). The plasma discharge with reactive ions commonly runs in a low-pressure vacuum vessel as reactor

lic nanostructures on III-V semiconductor substrates. For this purpose, the so-called *lift-off process* is preferred. The III-V wafer surface is covered with a positive resist (step 1). This type of resist is chemically modified by illumination, be it by light or by an electron beam, such that the illuminated areas become more soluble (breaking of polymer bonds). Subsequently the positive resist is illuminated by light or by electron beam radiation (step 2). After development, i.e. the removal of the resist in the illuminated areas (step 3) the wafer is covered by a structured resist film. The subsequent deposition of metal covers the resist and the open wafer areas equally (step 4). In the lift-off process the metal-covered resist structures are now removed in a solvent (step 5) and only the metal structures remain on the wafer, similar as in the direct metal etching process. Positive resists can also be used in order to transfer patterns directly into the semiconductor wafer. After the illumination of the resist (step 2) and removal of the illuminated areas (step 3) only one etching process (step 4), mostly by RIE, is used to generate the required pattern of holes or nanostructures extending from the wafer surface like a mesa (Spanish: table). In the last fifth step, the resist remaining on top of the mesas is removed. This structuring method is extremely useful in order to prepare three-dimensional quantum boxes (dots) based on semiconductor multiple heterostructures. An epitaxially deposited layer stack with two energetic barriers for electrons, e.g. of AlAs embedded in GaAs, creates a two-dimensional quantum confinement by means of the two energetic barriers (Sec. 12.7). Confinement in the third direction, i.e. formation of the quantum dot, is achieved by lithographic structuring of nanoscale mesa-like columns.

All structuring methods for nanostructures presented so far start with macroscopic solids and/or solid films and use lithographic techniques to machine the nanoscale structures into the solid (*top-down approach*). In recent years, so called bottom-up techniques have gained interest, in which the self-organization principle of nature, as we know it from crystal growth, is successfully applied. Under adequate physical and chemical conditions, nanostructures grow by self-organization without the use of complex and cost-intensive lithographic techniques. An example is the chemical self-organization of organic nanoparticles and fiber structures. Particularly in semiconductor physics, the self-organized epitaxial growth of semiconductor whiskers (wires, columns) has gained importance for the study of quantum transport and for future device applications. After deposition of metallic nanoparticles with dimensions of 10-100nm on a semiconductor wafer these particles act as crystallization nuclei and local catalysts in the subsequent epitaxial growth of whiskers both in MBE and in MOVPE (MOCVD, panel XVIII). In the MOVPE process, the metallic nuclei locally enhance the catalytic decomposition of the gaseous precursors and induce epitaxial growth under special growth conditions only on the nuclei. This leads to whisker growth; 10 to 100nm thick nanocolumns with lengths up to several micrometers grow. In the case of III-nitrides, whisker growth occurs

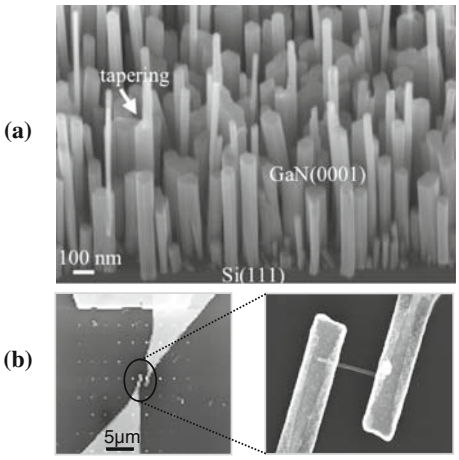

Fig. XVIII.3. Scanning electron microscopy image of an arrangement of GaN whiskers grown by molecular beam epitaxy on a Si(111) substrate (**a**) and of one single whisker which has been removed from the substrate and is placed on a new Si/SiO_2 host substrate (**b**). Electron beam lithographically structured markers (*bright spots*) were deposited in advance on the host substrate. The whisker is electrically contacted by metal contacts (large bright areas) that are also prepared by electron beam lithography. (After Calarco et al. [XVIII.2])

in MBE under nitrogen rich conditions even without the use of metallic nuclei (Fig. XVIII.3a). After removal of the nanocolumns from the substrate in an ultrasonic bath they can be spread on a high resistive host substrate (SiO_2 etc.), electrically contacted by lithography (Fig. XVIII.3b) and subsequently used for the study of quantum transport.

For eventual applications in quantum electronics, an ordered growth of semiconductor nanocolumns with respect to thickness, length and local arrangement is desirable. This can be achieved by depositing metal dots of homogeneous thickness on well-defined sites on the wafer. Well-ordered arrangement of nanocolumns can also be achieved by selective growth through mask holes, which are deposited on the wafer in advance. As an example, Fig. XVIII.4 shows ordered selective growth of GaAs columns with a diameter of about 150nm on GaAs(111). The mask con-

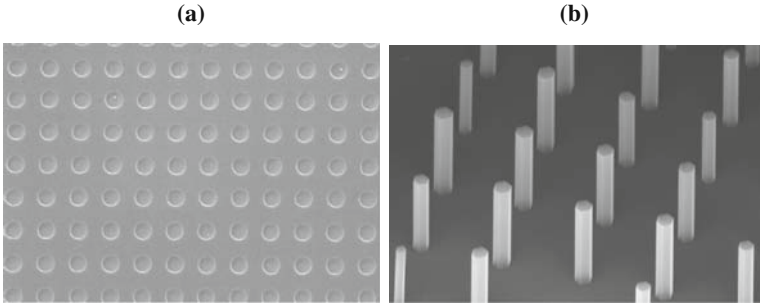

Fig. XVIII.4. Ordered selective growth of GaAs nanocolumns on a masked GaAs(111) substrate (scanning electron microscopic images). (**a**) HSQ (hydrogen silesquioxane) mask prepared on a GaAs(111) wafer by electron beam lithography; (**b**) GaAs nanocolumns (diameter approx. 180 nm) grown in metal-organic vapor phase epitaxy (MOVPE, panel XVII) by use of the HSQ mask in a). (After Klinger et al. [XVIII.3])

sisting of an inorganic negative resist (HSQ = Hydrogen Silesquioxane) was fabricated by electron beam lithography. The GaAs columns were grown in MOVPE at a growth temperature of 750°C using AsH_3 and TMG (Trimethyl Gallium) as precursors. Once more, it should be emphasized that the selective growth is essential as is possible preferentially in MOVPE, where the non-reactive surface of the mask does not catalytically dissociate the precursors and therefore does not allow GaAs growth. Growth of semiconductor columns was meanwhile demonstrated in MBE, MOVPE, and in MOCVD for the important semiconductors Si, SiGe, and GaAs, InAs, GaN, InN and ZnO. Hence, interesting new developments are expected that may lead to a semiconductor-based quantum electronics.

References

XVIII.1 M. Nishikino, M. Tanaka, K. Nagashima, M. Kishimoto, M. Kado, T. Kawachi, K. Sukegawa, Y. Ochi, N. Hasegawa, Y. Kato: Phys. Rev. A **68**, 061802 (2003)

XVIII.2 R. Calarco, M. Marso, R. Meijers, Th. Richter, N. Aykanat, T. Thillosen, T. Stoica, H. Lüth: NanoLetters **5** (2005) 981

XVIII.3 V. Klinger, J. Wensorra (Research Centre Jülich 2007): priv. commun.