Chapter 6
Band Structure

Silicon is a metal.

A.H. Wilson, 1931 [74]

Abstract A treatment of electron states in one-dimensional potentials introduces into the concepts of band gap and effective mass. The band structures of various semiconductors are reviewed. The systematics of band gaps, symmetry considerations, band gaps in alloys, amorphous semiconductors and the effect of strain and temperature are discussed. Electron and hole dispersions are treated and the density of states in various dimensions is derived.

6.1 Introduction

Valence electrons that move in the crystals feel a periodic potential

$$U(\mathbf{r}) = U(\mathbf{r} + \mathbf{R}) \tag{6.1}$$

for all vectors $\mathbf{R}$ of the direct lattice. The potential[1] is due to the effect of the ion cores and all other electrons. Thus a serious many-body problem is present. In principle, the band structure can be calculated from the periodic arrangements of the atoms and their atomic order number. We note that for some problems, e.g. the design of optimal solar cells, a certain band structure is known to be ideal and a periodic atomic arrangement, i.e. a material, needs to be found that generates the optimal band structure. This problem is called the *inverse band structure problem*.

6.2 Electrons in a Periodic Potential

6.2.1 Bloch's Theorem

First, we will deduce some general conclusions about the structure of the solution as a consequence of the periodicity of the potential. We first investigate the solution of a Schrödinger equation of the type

[1] In this book the form of the potential will never be explicitly given.

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
https://doi.org/10.1007/978-3-030-51569-0_6

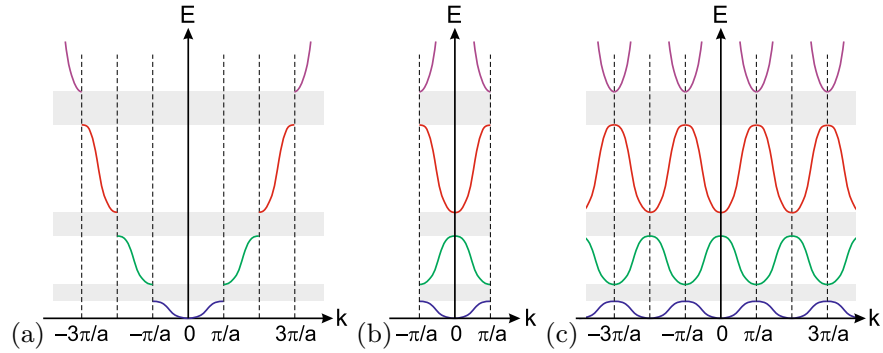

Fig. 6.1 Zone schemes for a band structure: **a** extended, **b** reduced and **c** repetitive zone scheme

$$H\,\Psi(\mathbf{r}) = \left[-\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r})\right]\Psi(\mathbf{r}) = E\,\Psi(\mathbf{r}) \tag{6.2}$$

for an electron. U will be periodic with the lattice, i.e. it will obey (6.1).

Bloch's theorem says that the eigenstates Ψ of a one-particle Hamiltonian as in (6.2) can be written as the product of plane waves and a lattice-periodic function, i.e.

$$\Psi_{n\mathbf{k}}(\mathbf{r}) = A\,\exp(\mathrm{i}\,\mathbf{k}\,\mathbf{r})\,u_{n\mathbf{k}}(\mathbf{r})\ . \tag{6.3}$$

The normalization constant A is often omitted. If $u_{n\mathbf{k}}(\mathbf{r})$ is normalized, $A = 1/\sqrt{V}$, where V is the integration volume. The wavefunction is indexed with a quantum number n and the wavevector $\mathbf{k}$. The key is that the function $u_{n\mathbf{k}}(\mathbf{r})$, the so-called Bloch function, is periodic with the lattice, i.e.

$$u_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}\,(\mathbf{r} + \mathbf{R}) \tag{6.4}$$

for all vectors $\mathbf{R}$ of the direct lattice. The proof is simple in one dimension and more involved in three dimensions with possibly degenerate wavefunctions, see [451].

If $E_{n\mathbf{k}}$ is an energy eigenvalue, then $E_{n\mathbf{k}+\mathbf{G}}$ is also an eigenvalue for all vectors $\mathbf{G}$ of the reciprocal lattice, i.e.

$$E_n(\mathbf{k}) = E_n\,(\mathbf{k} + \mathbf{G})\ . \tag{6.5}$$

Thus the energy values are periodic in reciprocal space. The proof is simple, since the wavefunction (for $\mathbf{k}+\mathbf{G}$) $\exp(\mathrm{i}(\mathbf{k}+\mathbf{G})\mathbf{r})u_{n(\mathbf{k}+\mathbf{G})}(\mathbf{r})$ is for $u_{n(\mathbf{k}+\mathbf{G})}(\mathbf{r}) = \exp(-\mathrm{i}\mathbf{G}\mathbf{r})u_{n\mathbf{k}}(\mathbf{r})$ obviously an eigenfunction to $\mathbf{k}$.

A band structure along one $\mathbf{k}$-direction can be displayed in various zone schemes as depicted in Fig. 6.1. The most frequently used scheme is the *reduced* zone scheme. In three dimensions, the band structure is typically shown along particular paths in the Brillouin zone, as depicted, e.g., in Fig. 6.2c.

6.2.2 Free-Electron Dispersion

If the entire wavefunction (from (6.3)) obeys the Schrödinger equation (6.2), the Bloch function $u_{n\mathbf{k}}$ fulfills the equation

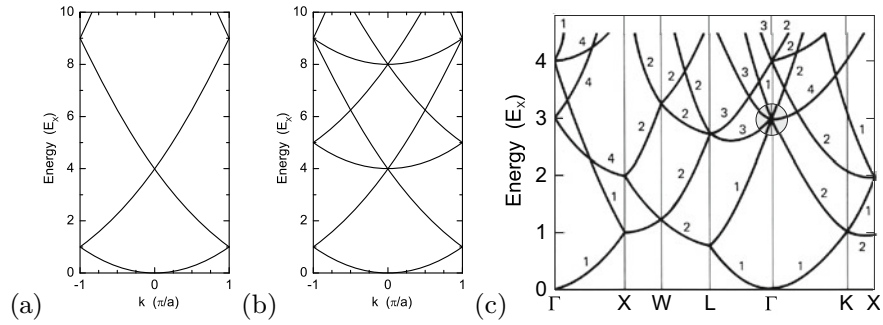

Fig. 6.2 Dispersion of free electrons (empty lattice calculation, $U = 0$, shown in the first Brillouin zone) in **a** a one-dimensional lattice ($\mathbf{G} = n\,2\pi/a$), **b** a simple cubic lattice ($\mathbf{G} = (h, k, l)\,2\pi/a$) and **c** in a fcc lattice. The energy is measured in units of the energy at the X-point, $E_X = (\hbar^2/2m)(\pi/a)^2$. The *shaded circle* in (**c**) represents the region where the band gap develops for finite periodic potential $U \neq 0$

$$\left[\frac{1}{2m}\,(\mathbf{p} + \hbar\,\mathbf{k})^2 + U(\mathbf{r})\right] u_{n\mathbf{k}}(\mathbf{r}) = E_{n\mathbf{k}}\,u_{n\mathbf{k}}(\mathbf{r})\ , \tag{6.6}$$

which is easy to see from $\mathbf{p} = -i\hbar\nabla$.

First, we discuss the simplest case of a periodic potential, $U \equiv 0$. This calculation is also called the *empty lattice* calculation. The solution of (6.6) is then just constant, i.e. $u_{\mathbf{k}} = c$ and $\Psi_{\mathbf{k}}(\mathbf{r}) = c\,\exp(\mathrm{i}\mathbf{k}\mathbf{r})$. The dispersion of the free electron is then given by

$$E(\mathbf{k}) = \frac{\hbar^2}{2m}\,\mathbf{k}^2\ , \tag{6.7}$$

where $\mathbf{k}$ is an arbitrary vector in the reciprocal space. $\mathbf{k}'$ is a vector from the Brillouin zone such that $\mathbf{k} = \mathbf{k}' + \mathbf{G}$ with a suitable reciprocal lattice vector $\mathbf{G}$. Because of (6.5) the dispersion relation can be written also as

$$E(\mathbf{k}) = \frac{\hbar^2}{2m}\,(\mathbf{k}' + \mathbf{G})^2\ , \tag{6.8}$$

where $\mathbf{k}'$ denotes a vector from the Brillouin zone. Thus, many branches of the dispersion relation arise from using various reciprocal lattice vectors in (6.8).

The resulting dispersion relation for the free electron is shown in Fig. 6.2a for a one-dimensional system ($\mathbf{k}'$ and $\mathbf{G}$ are parallel) and in Fig. 6.2b for the simple cubic lattice (in the so-called reduced zone scheme). In Fig. 6.2c, the (same) dispersion of the free electron is shown for the fcc lattice.

6.2.3 Non-Vanishing Potential

Now the effect of a non-vanishing periodic potential on electron motion will be discussed. A simple, analytically solvable model that visualizes the effect of a periodic potential on the dispersion relation of the electrons and the formation of a (one-dimensional) band structure with gaps is the Kronig-Penney model [71] which is discussed in the Appendix F.

6.2.3.1 General Wave Equation

In this section, we will discuss the solution of a general wave equation for electrons in a periodic potential. The solution is investigated particularly at the zone boundary. The potential U is periodic with the lattice (6.1). It can be represented as a Fourier series with the reciprocal lattice vectors (lattice vector expansion, cf. (3.19)):

$$U(\mathbf{r}) = \sum_{\mathbf{G}} U_{\mathbf{G}} \, \exp(\mathrm{i}\,\mathbf{G}\,\mathbf{r}) \ . \tag{6.9}$$

Since U is a real function, $U_{-\mathbf{G}} = U_{\mathbf{G}}^*$. The deeper reason for the success of such an approach is that for typical crystal potentials, the Fourier coefficients decrease rapidly with increasing $\mathbf{G}$, e.g. for the unscreened Coulomb potential $U_{\mathbf{G}} \propto 1/G^2$. The wavefunction is expressed as a Fourier series (or integral) over all allowed (Bloch) wavevectors $\mathbf{K}$,

$$\Psi(\mathbf{r}) = \sum_{\mathbf{K}} C_{\mathbf{K}} \, \exp(\mathrm{i}\,\mathbf{K}\,\mathbf{r}) \ . \tag{6.10}$$

The kinetic and potential energy terms in the Schrödinger equation (6.6) are

$$\nabla^2 \Psi = -\sum_{\mathbf{K}} \mathbf{K}^2 \, C_{\mathbf{K}} \, \exp(\mathrm{i}\,\mathbf{K}\,\mathbf{r}) \tag{6.11a}$$

$$U\Psi = \sum_{\mathbf{G}} \sum_{\mathbf{K}} U_{\mathbf{G}} \, C_{\mathbf{K}} \, \exp(\mathrm{i}\,(\mathbf{G}+\mathbf{K})\,\mathbf{r}) \ . \tag{6.11b}$$

With $\mathbf{K}' = \mathbf{K} + \mathbf{G}$, (6.11b) can be rewritten as

$$U\Psi = \sum_{\mathbf{G}} \sum_{\mathbf{K}'} U_{\mathbf{G}} \, C_{\mathbf{K}'-\mathbf{G}} \, \exp\left(\mathrm{i}\,\mathbf{K}'\,\mathbf{r}\right) \ . \tag{6.12}$$

Now, the Schrödinger equation can be written as an (infinite) system of algebraic equations:

$$(\lambda_{\mathbf{K}} - E)\, C_{\mathbf{K}} + \sum_{\mathbf{G}} U_{\mathbf{G}} \, C_{\mathbf{K}-\mathbf{G}} = 0 \ , \tag{6.13}$$

with $\lambda_{\mathbf{K}} = \hbar^2 \, \mathbf{K}^2/(2m)$.

6.2.3.2 Solution for One Fourier Coefficient

The simplest (non-trivial) potential energy has only one important Fourier coefficient $-U$ ($U > 0$) for the shortest reciprocal lattice vector $\mathbf{G}$. Also, we have $U_{-\mathbf{G}} = U_{\mathbf{G}}$. Thus, the (one-dimensional) potential has the form $U(x) = -2\,U\,\cos(Gx)$. Then the equation system (6.13) has only two equations for $C_{\mathbf{K}}$ and $C_{\mathbf{K}-\mathbf{G}}$, leading to the condition

$$\begin{vmatrix} \lambda_{\mathbf{K}} - E & -U \\ -U & \lambda_{\mathbf{K}-\mathbf{G}} - E \end{vmatrix} = 0 \ . \tag{6.14}$$

We find two solutions

$$E_{\pm} = \frac{\lambda_{\mathbf{K}} + \lambda_{\mathbf{K}-\mathbf{G}}}{2} \pm \sqrt{\left(\frac{\lambda_{\mathbf{K}} - \lambda_{\mathbf{K}-\mathbf{G}}}{2}\right)^2 + U^2} \ . \tag{6.15}$$

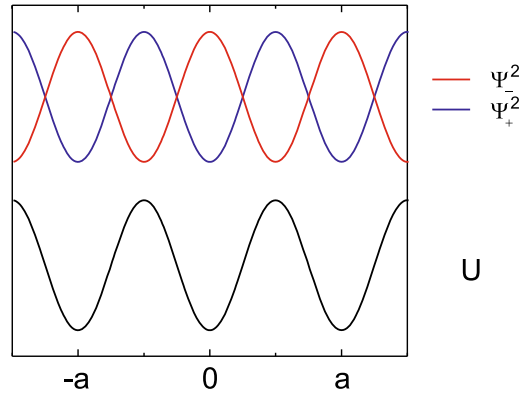

Fig. 6.3 Periodic potential U (one-dimensional cosine, *black*) and the squares of the wavefunctions Ψ_- (*red*) and Ψ_+ (*blue*) for the wavevector at the zone boundary, $K = G/2 = \pi/a$

6.2.3.3 Solution at the Zone Boundary

We consider the solution at the zone boundary, i.e. at $\mathbf{K} = \mathbf{G}/2$. The kinetic energy is then the same for $\mathbf{K} = \pm\mathbf{G}/2$, i.e. $\lambda_{\mathbf{K}} = \lambda_{\mathbf{K}-\mathbf{G}} = (\hbar^2/2m)\,(G^2/4) = \lambda$. The determinant (6.14) reads then

$$(\lambda - E)^2 - U^2 = 0\,. \tag{6.16}$$

Thus the energy values at the zone boundary are

$$E_\pm = \lambda \pm U = \frac{\hbar^2}{2m}\,\frac{G^2}{4} \pm U\,. \tag{6.17}$$

At the zone boundary, a splitting of the size $E_+ - E_- = 2U$ occurs. The center of the energy gap is given by the energy $\lambda_{\mathbf{K}}$ of the free-electron dispersion. The ratio of the coefficients is $C_{\mathbf{G}/2}/C_{-\mathbf{G}/2} = \mp 1$. The '$-$' solution of (6.17) (lower energy) is a standing cosine wave (Ψ_-), the '$+$' solution (Ψ_+) is a standing sine wave as visualized in Fig. 6.3. For the lower-energy (binding) state the electrons are localized at the potential minima, i.e. at the atoms, for the upper state (antibinding) the electrons are localized between the atoms. Both wavefunctions have the same periodicity since they belong to the same wavevector $\mathbf{K} = \mathbf{G}/2$. We note that the periodicity of Ψ is $2a$, while the periodicity of Ψ^2 is equal to the lattice constant a.

6.2.3.4 Gap States

For energies within the gap, solutions with a *complex* wavevector $K = G/2 + \mathrm{i}\,q$ exist. Solving (6.16) results (in terms of $q'^2 = (\hbar^2/2m)\,q^2$) to

$$E_\pm = \lambda - q'^2 \pm \sqrt{-4\,\lambda\,q'^2 + U^2}\,. \tag{6.18}$$

For energies $E = \lambda + \epsilon$ with $-U \le \epsilon \le U$, the complex part of the wavevector is given by

$$q'^2 = -(\epsilon + 2\lambda) + \sqrt{4\lambda\,(\epsilon + \lambda) + U^2}\,. \tag{6.19}$$

The maximum value of q is in the center of the band gap ($\epsilon = 0$); for $|U| \ll 2\,\lambda$, it is $q'^2_{\max} \approx U^2/(4\lambda)$. At the band edges ($\epsilon = \pm U$), $q = 0$. q is the characteristic length of an exponentially decaying wave

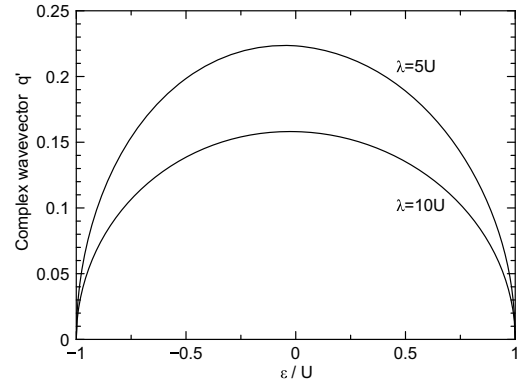

Fig. 6.4 Complex band structure $q'(\epsilon)$ according to (6.19) for two different values of λ/U

function. Such solutions occur at surfaces or interfaces. For larger band gaps, the localization length is smaller (larger q) (Fig. 6.4).

6.2.3.5 Solution in the Vicinity of the Zone Boundary

For $\mathbf{K}$ in the vicinity of the zone boundary the solutions (6.15) can be developed. Therefore, we use the (small) distance from the zone boundary $\widetilde{\mathbf{K}} = \mathbf{K} - \mathbf{G}/2$. With $\lambda = (\hbar^2/2m)\,(G^2/4)$ we rewrite still exactly (6.15):

$$E_{\pm}\left(\widetilde{\mathbf{K}}\right) = \frac{\hbar^2}{2m}\left(\frac{1}{4}\mathbf{G}^2 + \widetilde{\mathbf{K}}^2\right) \pm \left(4\lambda\,\frac{\hbar^2\widetilde{\mathbf{K}}^2}{2m} + U^2\right)^{1/2} . \tag{6.20}$$

For small $\widetilde{\mathbf{K}}$ with $\frac{\hbar^2\mathbf{G}\widetilde{\mathbf{K}}}{2m} \ll |U|$, the energy is then approximately given by

$$E_{\pm}\left(\widetilde{\mathbf{K}}\right) \cong \lambda \pm U + \frac{\hbar^2\widetilde{\mathbf{K}}^2}{2m}\left(1 \pm \frac{2\,\lambda}{U}\right) . \tag{6.21}$$

Thus the energy dispersion in the vicinity of the zone boundary is parabolic. The lower state has a negative curvature, the upper state a positive curvature. The curvature is

$$m^* = m\,\frac{1}{1 \pm 2\lambda/U} \approx \pm m\,\frac{U}{2\lambda} , \tag{6.22}$$

and will be later related to the *effective mass*. The approximation in (6.22) is valid for $|U| \ll 2\lambda$. We note that in our simple model m^* increases linearly with increasing band gap $2U$ (see Fig. 6.34 for experimental data).

6.2.4 Kramer's Degeneracy

$E_{\mathrm{n}}(\mathbf{k})$ is the dispersion in a band. The time-reversal symmetry (Kramer's degeneracy) implies

$$E_{\mathrm{n}\uparrow}(\mathbf{k}) = E_{\mathrm{n}\downarrow}(-\mathbf{k}) , \tag{6.23}$$

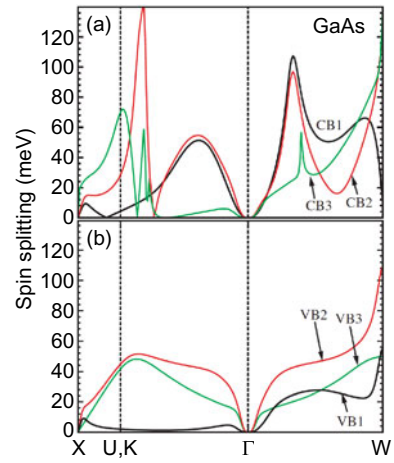

Fig. 6.5 Theoretical calculation of the spin splitting of **a** the three lowest conduction bands (CB1, CB2, and CB3) and **b** the top three valence bands (VB1, VB2, and VB3) of GaAs. Adapted from [453]

where the arrow refers to the direction of the electron spin. If the crystal is symmetric under inversion, we have additionally

$$E_{\mathrm{n}\uparrow}(\mathbf{k}) = E_{\mathrm{n}\uparrow}(-\mathbf{k}) \ . \tag{6.24}$$

With both time reversal and inversion symmetry the band structure fulfills

$$E_{\mathrm{n}\uparrow}(\mathbf{k}) = E_{\mathrm{n}\downarrow}(\mathbf{k}) \ . \tag{6.25}$$

The inversion symmetry is particularly important for the spin-orbit interaction. In the absence of inversion symmetry, e.g. in (non-centrosymmetric) zincblende crystals (Fig. 3.16b) or in heterostructures (Fig. 12.35b), a spin splitting, e.g. $E_{\mathrm{n}\uparrow}(\mathbf{k}) \neq E_{\mathrm{n}\downarrow}(\mathbf{k})$, is present. It can be thought of as provoked by an effective magnetic field. Bulk inversion asymmetry (BIA) leads to the Dresselhaus spin splitting [452, 453] that is shown in Fig. 6.5 for GaAs (cmp. Fig. 6.10a). The spin splitting due to structural inversion asymmetry (SIA) is described by the Bychkov-Rashba Hamiltonian [454, 455]. A review on these topics can be found in [456].

6.2.5 Symmetry Considerations

In general the symmetry of the lattice is a symmetry of the system's Hamiltonian and thus transfers into the electronic (and other) properties of the semiconductor. The means to formulate this mathematically is group theory and representation theory. At a given reciprocal lattice point, the wave function must fulfill the given spatial symmetry. Additional symmetry due to spin and spin-orbit interaction enters via the double-group scheme. This problem has been treated for the 32 point groups (cmp. Table B.2) in [457] and in [458] particularly for the pc, fcc, bcc and hcp lattices. A detailed treatment for the zincblende [459] and wurtzite [460] structures have been given. The most popular Hamiltonians are treated in [461].

The symmetry at particular points in direct or reciprocal space is denoted by the irreducible representations of the symmetry (point) group, e.g. by the Γ_i-symbols used in Figs. 6.9, 6.10 or also Fig. 6.44. As an example, base functions with the symmetry of the irreducible representations of tetraeder group $\mathrm{T_d}$ are listed in Table 6.1. With the knowledge of the wave functions at the points of high symmetry, it is possible to deduce the general nature of the energy bands in the vicinity of such symmetry points.

Table 6.1 Representations of the tetraeder group (zincblende structure) in molecular, BSW [462] and Koster [457] notation and (examples of) corresponding base functions (c.p.: cyclic permutations)

molecular	BSW	Koster	base functions
A_1	Γ_1	Γ_1	$x\,y\,z$, $x^2+y^2+z^2$
A_2	Γ_2	Γ_2	$x^4\,(y^2-z^2)$ + c.p.
E	Γ_{12}	Γ_3	$2\,z^2-(x^2+y^2)$, (x^2-y^2)
T_2	Γ_{15}	Γ_4	x, y, z, xy, xz, yz
T_1	Γ_{25}	Γ_5	$z\,(x^2-y^2)$ and c.p.

6.2.6 *Topological Considerations*

Starting with research on the quantum Hall effect and based on previous mathematical theorems, it has become clear that the band structure of 'insulators' has topological properties which in turn lead to a elegant classification of materials (and many effects/phases) [370, 463, 464]. In this context, the term 'insulator' means a material with gap between filled and empty states, i.e. semiconductors are exactly like this if the temperature is not too high (related to the gap divided by k_B). We recall the discussion of the diatomic linear chain in Sect. 5.2.3 where the bands turned out to have different topological properties depending on the ratio of sporing constants.

Topology is a branch of mathematics where objects that are related to each other by a smooth deformation are classified as the same. For example, a sphere and an ellipsoid are topologically the same. Also, a doughnut and a cup are the same since they have one hole. A quantity that is independent of such smooth transformations is termed 'topological invariant'. Such a number is the *genus* g of a surface that counts the number of holes. According to the Gauss-Bonnet theorem, the integral of the Gaussian curvature K over a closed surface S is given by

$$\int_S K \,\mathrm{d}A = 2\pi\,(2-2\,g)\;. \tag{6.26}$$

The Gaussian curvature $K = \kappa_1\,\kappa_2$ of a (differentiable) surface in 3D is the product of the principal curvatures κ_1 and κ_2 (maximum and minimum curvature of the curves from all normal planes that contain the normal vector intersecting with the surface). For a sphere of radius r, the Gaussian curvature is $1/r^2$ everywhere and the integral in (6.26) is $4\,\pi$, making $g = 0$. For a topologically different example we look at a torus (all points that have the fixed distance r from a circle of radius R, $r < R$). It is parametrized by

$$\boldsymbol{r} = R\begin{pmatrix}\cos\phi\\ \sin\phi\\ 0\end{pmatrix} + r\begin{pmatrix}\cos\phi\,\cos\theta\\ \sin\phi\,\cos\theta\\ \sin\theta\end{pmatrix}\;, \tag{6.27}$$

with both the angles ϕ and θ running between 0 and $2\,\pi$. The principal curvature κ_1 along the θ-direction is $1/r$ (for all ϕ). The other principal curvature κ_2 in azimuthal (ϕ) direction changes sign with θ (positive outside, negative inside) and is also independent of ϕ. Its integral over the outer and inner part cancel exactly, thus the integral of $\kappa_1\,\kappa_2$ over the entire torus is zero and therefore $g = 1$.

Next, we connect the periodicity of the Brillouin zone in two dimensions with variables on a torus in 3D as shown in Fig. 6.6 (cmp. Fig. 5.3 for the 1D case). This concept can be generalized for a 3D band structure and a torus in four dimensions.

If a constant function $f = n\,a\,b/(2\,\pi)$ is integrated over the Brillouin zone (X is at $\pm\pi/a$, Y is at $\pm\pi/b$), the integral is

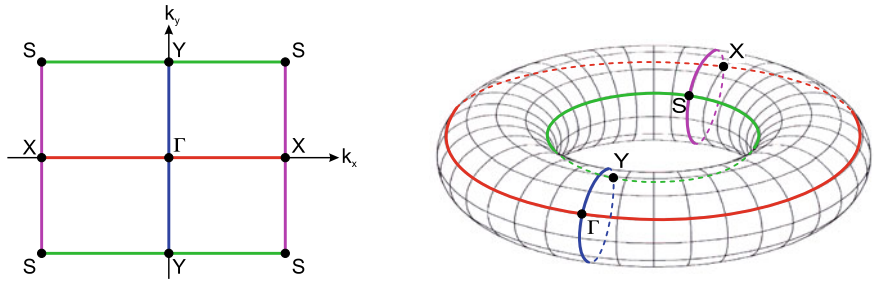

Fig. 6.6 Brillouin zone of a two-dimensional rectangular lattice and mapping to a torus

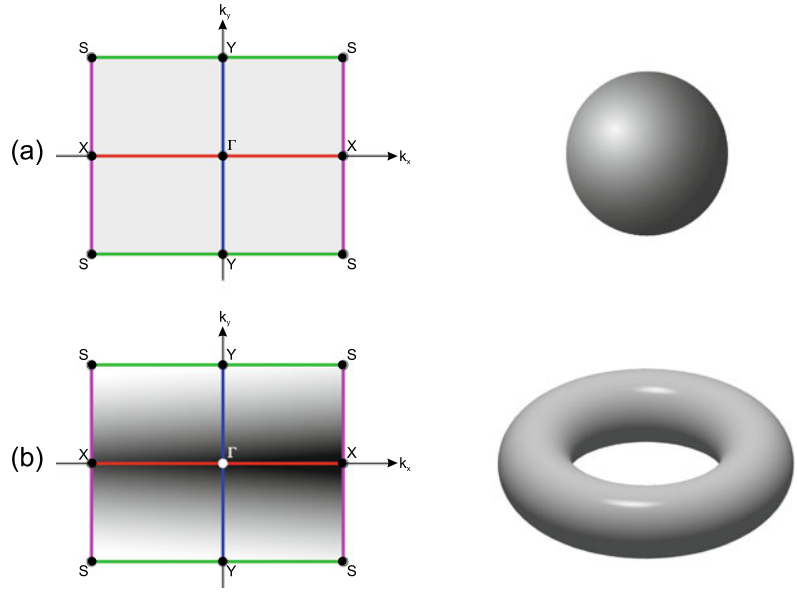

Fig. 6.7 Brillouin zone of a rectangular lattice with **a** a constant function and **b** a function that changes sign. On the right the topology of the situation is visualized

$$\frac{1}{2\pi}\int_{\mathrm{BZ}} f(\boldsymbol{k})\,\mathrm{d}^2\boldsymbol{k} = n\ . \tag{6.28}$$

If for example another function f that is independent of k_x and changes sign in k_y-direction with $\int f\,\mathrm{d}k_y = 0$, similar to the curvature of the torus, is integrated over the Brillouin zone, the results will be zero. This is schematically shown in Fig. 6.7 if the integrand is interpreted as curvature.

The generalization of the Berry phase [369] to Bloch states has been made in [465, 466]. For a two-dimensional system with Bloch bands and with Bloch functions $u_m(\boldsymbol{k})$ as in (6.3), the integrand leading to a topological invariant is given by the Berry connection (cmp. (5.29)) $\mathcal{A}_m = \langle u_m(\boldsymbol{k})|\imath\,\nabla_{\boldsymbol{k}}|u_m(\boldsymbol{k})\rangle$ and its Berry curvature or Berry flux in three-dimensional notation $\mathcal{F}_m = \nabla_{\boldsymbol{k}} \times \mathcal{A}_m$. The Chern number C_m for a band (separated by gaps from other bands), defined as integral over the Brillouin zone,

$$C_m = \frac{1}{2\pi}\int_{\mathrm{BZ}} \mathcal{F}_m\,\mathrm{d}^2\boldsymbol{k}\ , \tag{6.29}$$

takes only integer values and is a topological invariant. That means that small variations of the Hamilton operator behind the band structure do not change its value. In the case of degeneracies, still the sum of Chern numbers over all occupied bands, $n = \sum_m n_m$, is a topological invariant as long as the empty

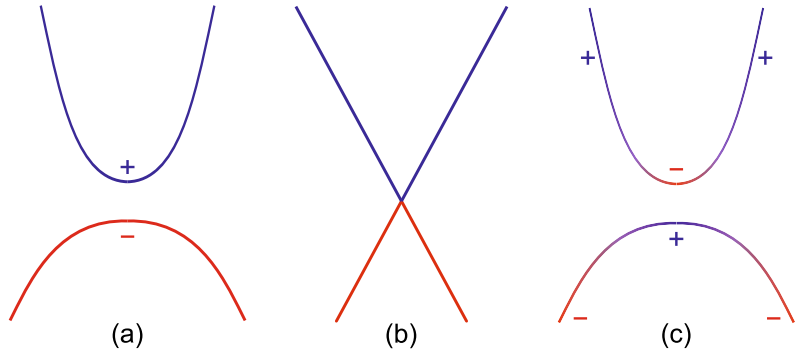

Fig. 6.8 Schematic band structures with (**a**) topologically trivial wave functions and **c** topologically non-trivial wave functions where within one band the character changes from s (*blue*, '+' for positive parity) to p (*red*, '−' for negative parity). In **b** the crossing point is visualized

states are separated by a gap. For a three-dimensional crystal and $\boldsymbol{k}$-space several topological invariants exist but a Chern number can be assigned to the Fermi surface or surface states (cmp. Sect. 11.6.3).

As an schematic example we show Fig. 6.8 for a topological trivial and non-trivial band structure. In the trivial bandstructure (as most semiconductors), the phase character of the wavefunction changes only little within the Brillouin zone, mostly p-type (negative parity) for the valence band and mostly s-type (positive parity) for the conduction band (Fig. 6.8a). In the topologically non-trivial bandstructure, band inversion takes place and the character of the wave function changes within a band (Fig. 6.8c). This sketch should be compared to Fig. 5.8 where a similar situation had been discussed for the lattice vibrations. An example for a semiconductor with band inversion is HgTe while CdTe or MnTe have trivial topology. Alloying leads at the transition from trivial to non-trivial (Fig. 6.8b) to zero-gap semiconductors (cf. Sect. 6.11).

6.3 Band Structures of Selected Semiconductors

In the following, the band structures of various important and prototype semiconductors are discussed. The band below the energy gap is called the valence band; the band above the gap is the conduction band. The band gap ΔE_{cv}, mostly denoted as E_{g}, is the energy separation between the highest valence-band state and the lowest conduction-band state. The maximum of the valence band is for most semiconductors at the Γ point.

6.3.1 Silicon

For silicon, an elemental semiconductor, (Fig. 6.9a) the minimum of the conduction band is located close to the X-point at $0.85\pi/a$ in the $\langle 100\rangle$ direction. Thus, it is not at the same point in **k** space as the top of the valence band. Such a band structure is called *indirect*. Since there are six equivalent $\langle 100\rangle$ directions, there are six equivalent minima of the conduction band.

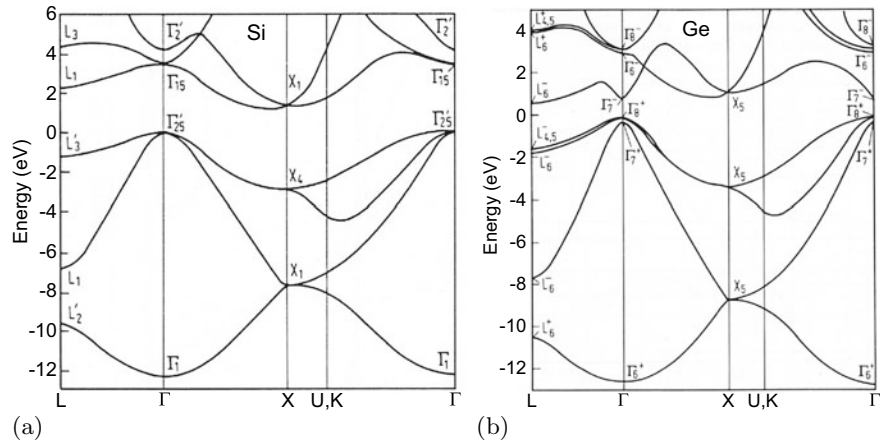

Fig. 6.9 Band structure of **a** silicon (indirect) and **b** germanium (indirect). In Si, the minima of the conduction band are in the ⟨100⟩ direction, for germanium in the ⟨111⟩ direction. Adapted from [164], based on [467]

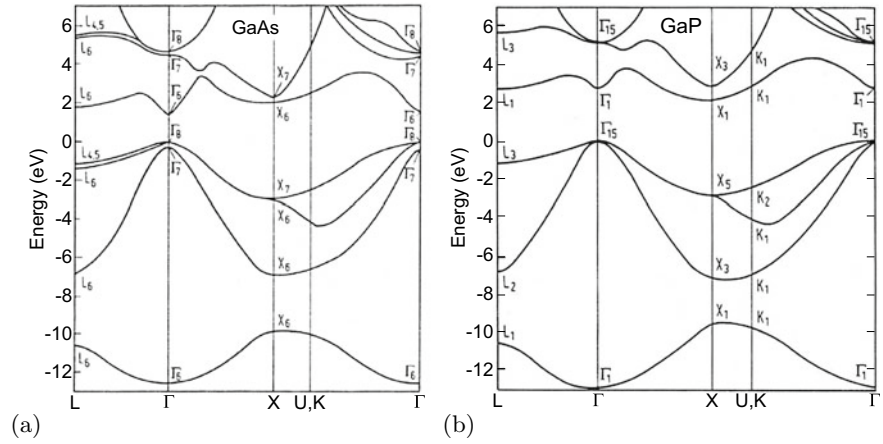

Fig. 6.10 Band structure of **a** GaAs (direct) and **b** GaP (indirect). For GaAs the minimum of the conduction band is at Γ, for GaP in the ⟨100⟩ direction. Adapted from [164], based on [467]

6.3.2 Germanium

Germanium, another elemental semiconductor, (Fig. 6.9b) also has an indirect band structure. The conduction minima are at the L point in the ⟨111⟩ direction. Due to symmetry there are eight equivalent conduction-band minima.

6.3.3 GaAs

GaAs (Fig. 6.10a) is a compound semiconductor with a *direct* band gap since the top of the valence band and the bottom of the conduction band are at the same position in **k** space (at the Γ-point). The next highest (local) minimum in the conduction band is close to the L point.

6.3.4 GaP

GaP (Fig. 6.10b) is an indirect compound semiconductor. The conduction-band minima are along the ⟨100⟩ directions.

6.3.5 GaN

GaN (Fig. 6.11) is a direct semiconductor that has wurtzite structure but can also occur in the metastable cubic (zincblende) phase.

6.3.6 Lead Salts

The band gap of PbS (Fig. 6.12), PbSe and PbTe is direct and located at the L point. The lead chalcogenide system shows the anomaly that with increasing atomic weight the band gap does not decrease monotonically. At 300 K, the band gaps are 0.41, 0.27 and 0.31 eV for PbS, PbSe and PbTe, respectively.

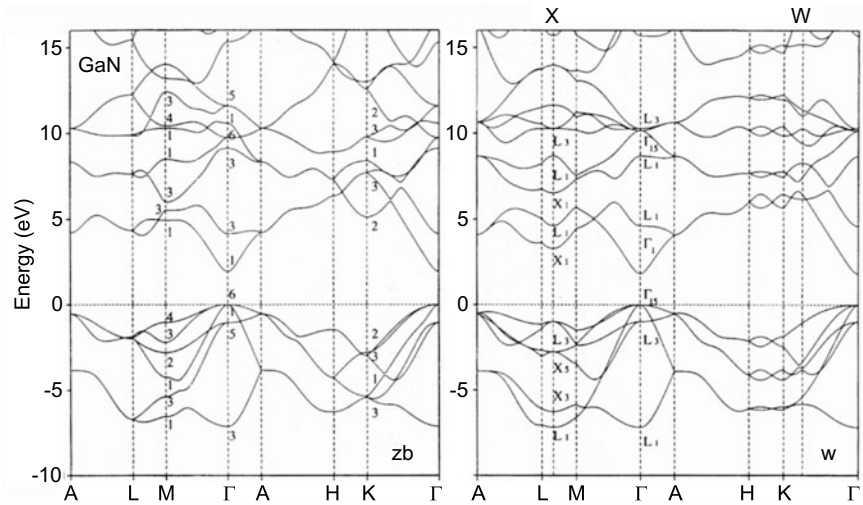

Fig. 6.11 Band structure of GaN (direct) in zincblende (zb) modification (*left*) and wurtzite (w) modification (*right*), both displayed in the wurtzite Brillouin zone to facilitate comparison

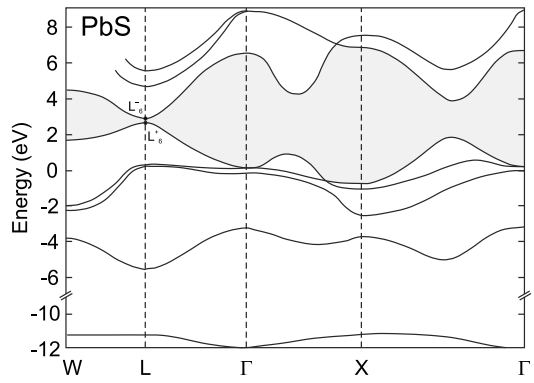

Fig. 6.12 Calculated band structure of PbS (direct). The energy gap is at the L point. The forbidden band is shown in *grey*. Adapted from [468]

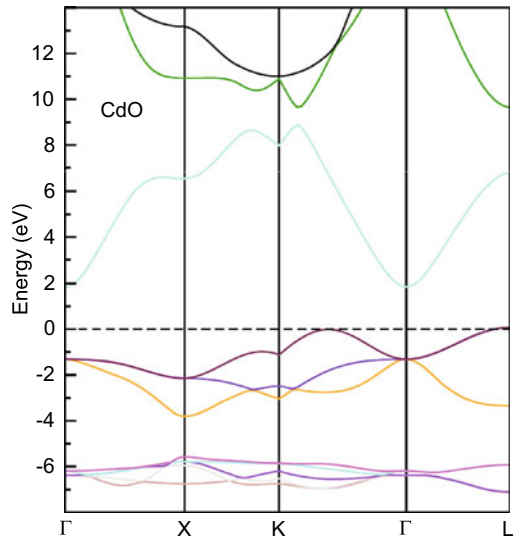

Fig. 6.13 Calculated indirect band structure of CdO. The top of the valence band is at $E = 0$. Adapted from [470]

6.3.7 MgO, ZnO, CdO

Cadmium oxide is a cubic semiconductor in the rocksalt structure. Due to symmetry considerations, coupling (repulsion) of oxygen 2p- and cadmium 3d-orbitals does not occur at the zone center in the rocksalt structure. Repulsion occurs though away from the Γ-point and therefore the valence band maximum is not located at the zone center (Fig. 6.13). Thus CdO is an indirect semiconductor. A similar effect would occur in rs-ZnO due to zinc 3d orbitals; however, ZnO has wurtzite structure for which p–d coupling at the Γ-point is allowed; thus ZnO is direct. In MgO, Mg of course only possesses populated s- and p-orbitals and no such repulsion is present; thus MgO even with its rocksalt structure is also direct [469].

6.3.8 Chalcopyrites

The experimental band gaps of a number of chalcopyrite semiconductors are listed in Table 6.2. The band structures of $CuAlS_2$, $CuAlSe_2$, and $CuGaSe_2$ are compared in Fig. 6.14.

In Fig. 6.15, the theoretical band structure of GaN and its closest related chalcopyrite $ZnGeN_2$ are compared, both shown in the chalcopyrite (orthorhombic) Brillouin zone. The band gap of $ZnGeN_2$ is smaller than that of GaN and the difference of 0.4 eV is fairly well reproduced by the calculation[2] (giving 0.5 eV).

[2]Due to the local density approximation (LDA) the absolute values of the band gaps are too small by about 1 eV.

Table 6.2 Band gaps of various chalcopyrite semiconductors

Material	E_g (eV)		E_g (eV)		E_g (eV)
$CuAlS_2$	3.5	$CuGaS_2$	2.5	$CuInS_2$	1.53
$CuAlSe_2$	2.71	$CuGaSe_2$	1.7	$CuInSe_2$	1.0
$CuAlTe_2$	2.06	$CuGaTe_2$	1.23	$CuInTe_2$	1.0–1.15
$AgAlS_2$	3.13	$AgGaS_2$	2.55	$AgInS_2$	1.87
$AgAlSe_2$	2.55	$AgGaSe_2$	1.83	$AgInSe_2$	1.24
$AgAlTe_2$	2.2	$AgGaTe_2$	1.1–1.3	$AgInTe_2$	1.0
$ZnSiP_2$	2.96	$ZnGeP_2$	2.34	$ZnSnP_2$	1.66
$ZnSiAs_2$	2.12	$ZnGeAs_2$	1.15	$ZnSnAs_2$	0.73
$CdSiP_2$	2.45	$CdGeP_2$	1.72	$CdSnP_2$	1.17
$CdSiAs_2$	1.55	$CdGeAs_2$	0.57	$CdSnAs_2$	0.26

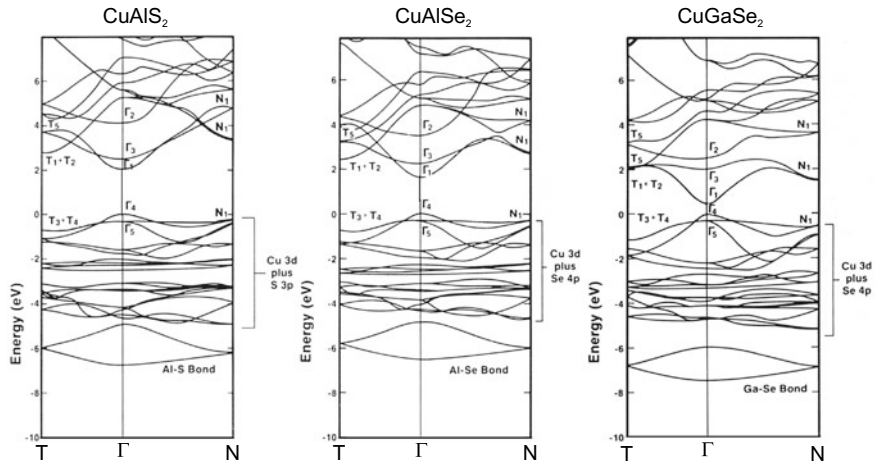

Fig. 6.14 Calculated band structures of $CuAlS_2$, $CuAlSe_2$, and $CuGaSe_2$. The absolute values of the gap energies are incorrect due to LDA calculation. Adapted from [471]

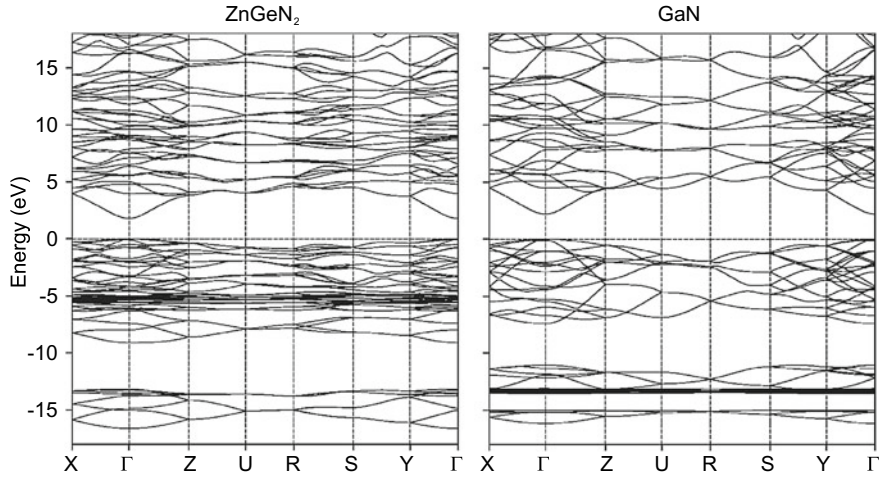

Fig. 6.15 Calculated (within LDA) band structures of $ZnGeN_2$ and its related III–V compound GaN, both displayed in the chalcopyrite (orthorhombic) Brillouin zone to facilitate comparison. Adapted from [472]

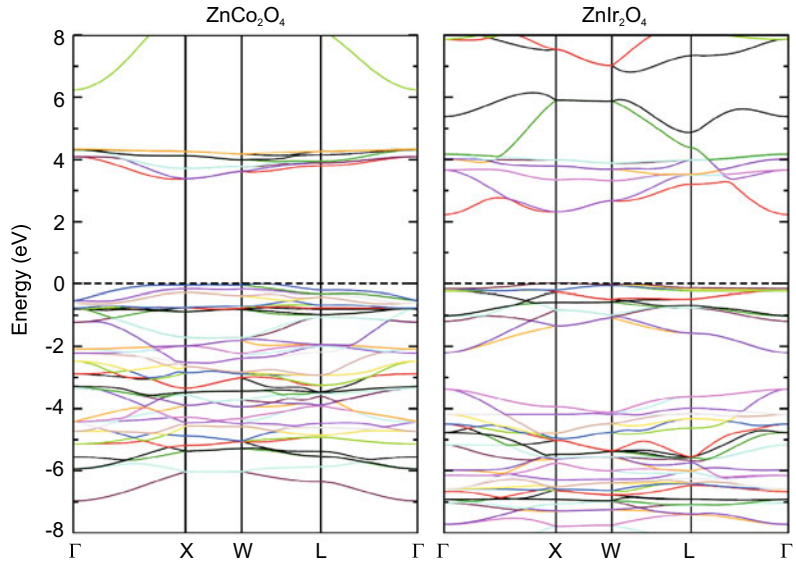

Fig. 6.16 Calculated band structures of $ZnCo_2O_4$ and $ZnIr_2O_4$. Adapted from [470]

6.3.9 Spinels

The band structure of spinels (in particular $CdIn_2S_4$) has been discussed in [473], the band structure of ZnM_2O_4 has been calculated in [470] for (M = Co, Rh, Ir) (Fig. 6.16) and in [474] for (M = Al, Ga, In).

6.3.10 Delafossites

In Fig. 6.17, the theoretical band structures of the delafossites $CuAlO_2$, $CuGaO_2$, and $CuInO_2$ are shown. The maximum of the valence band is not at Γ but near the F point. The direct band gap at Γ decreases for the sequence Al → Ga → In, similar to the trend for AlAs, GaAs and InAs. The direct band gap at F and L, causing the optical absorption edge, increases, however (experimental values are 3.5, 3.6, and 3.9 eV).

6.3.11 Perovskites

The calculated band structure of $BaTiO_3$ in the tetragonal phase is shown in Fig. 6.18. The minimum of the conduction band is at the Γ-point. The maximum of the valence band is not at the Γ-point but at the M point. The band gap of the LDA[3] calculation is too small (2.2 eV) compared to the experimental value ∼ 3.2 eV.

The band structure of the halide perovskites has been calculated for hybrid organic-inorganic compounds like $MAPbI_3$ and $FAPbI_3$ [476] and fully inorganic compounds $APbI_3$ (A = Li, Na, K, Rb, and Cs) [477]. Density of states and energy positions of $(MA,FA,Cs)(Pb,Sn)(Cl,Br,I)_3$ compounds are

[3]local density approximation.

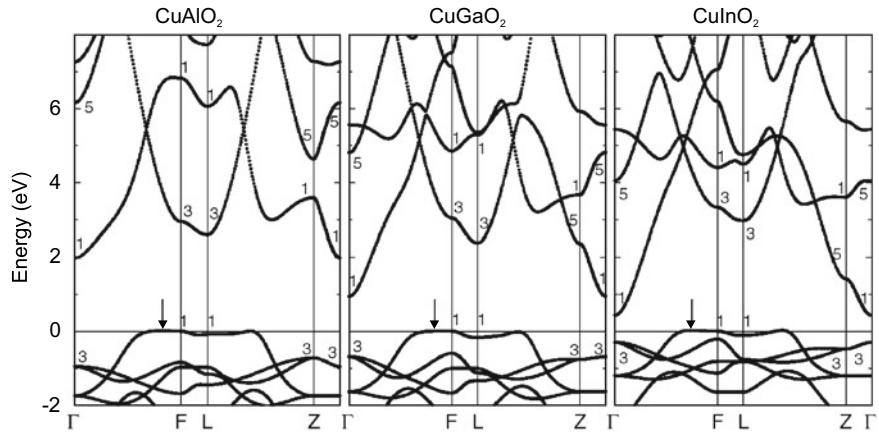

Fig. 6.17 Band structures of $CuAlO_2$, $CuGaO_2$, and $CuInO_2$, calculated with LDA (underestimating the absolute value of the band gaps). The *arrows* denote the maximum of the valence band that has been set to zero energy for each material. Adapted from [226]

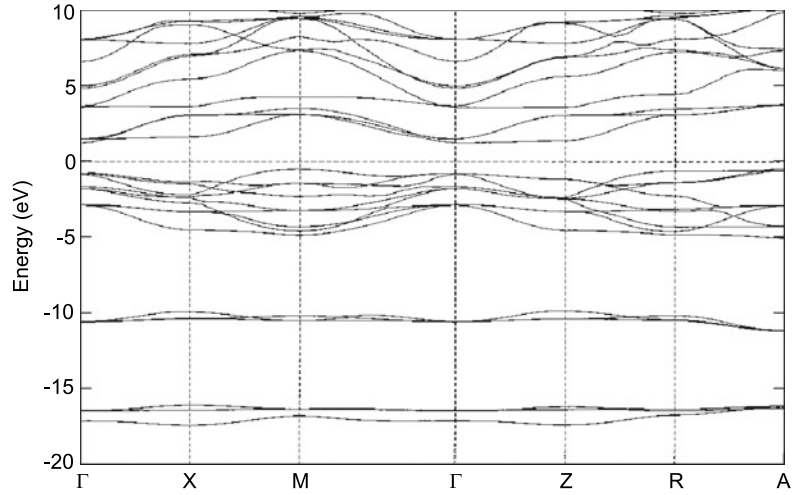

Fig. 6.18 Calculated energy band structure of $BaTiO_3$ along the major symmetry directions. The Fermi level (E_F) is set at zero energy. Adapted from [475]

discussed in [478]. The trends are summarized in Fig. 6.19. The band gap of halide perovskites can be varied across the visible range, e.g. within the $CsPb(Cl,Br,I)_3$ the system (Fig. 6.20).

6.4 Systematics of Semiconductor Band Gaps

The trends with regard to the size of the band gap for elemental, III–V and II–VI semiconductors can essentially be understood in terms of the bond strength and ionicity. In Fig. 6.21, the band gaps of many important semiconductors are shown as a function of the lattice constant. For elemental semiconductors, the band gap decreases with reduced bond strength, i.e. lattice constant (C→Si→Ge). A similar trend exists both for the III–V and the II–VI semiconductors.

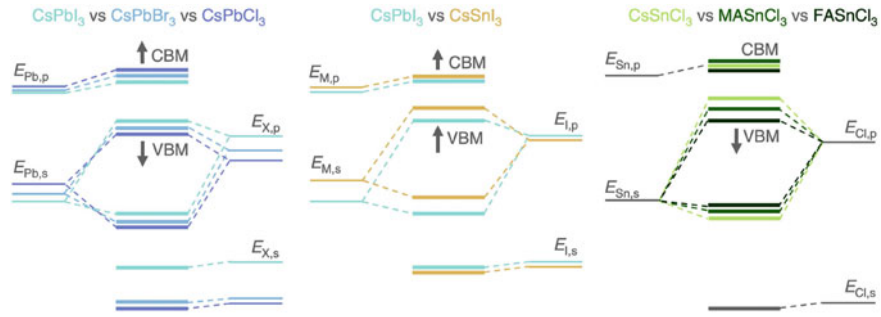

Fig. 6.19 Schematic energy levels in ABX_3 perovskites. Arrows indicate the shift of energy levels upon substitution of atoms or small organic molecules. Adapted from [478]

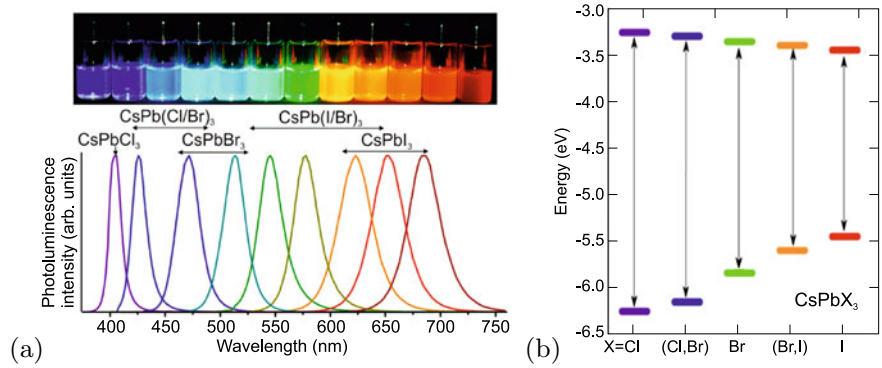

Fig. 6.20 **a** Photoluminescence of $CsPbX_3$ nanocrystals of halide perovskites with various anions (colloidal solutions in toluene under UV lamp ($\lambda = 365\,\text{nm}$)). Adapted from [479]. **b** Position of conduction and valence band edges of $CsPbX_3$ halide perovskites (relative to vacuum level at zero). Adapted from [480]

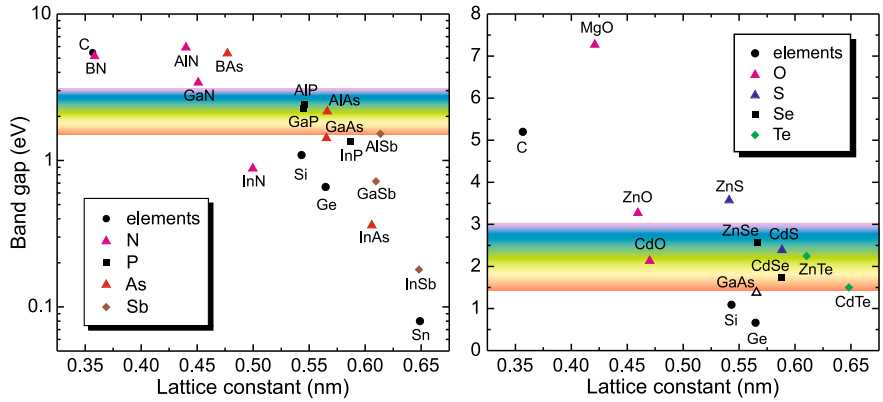

Fig. 6.21 Band gaps as a function of the lattice constant for various elemental, III–V and II–VI semiconductors. The lattice constant of wurtzite semiconductors has been recalculated for a cubic cell ($a^3_{\text{cubic}} = \sqrt{3}\,a^2\,c$)

For the same lattice constant, the band gap increases with increasing ionicity, i.e. IV–IV→ III–V→II–VI→I–VII. A typical example is the sequence Ge→ GaAs→ZnSe→ CuBr, for which all materials have almost the same lattice constant, and the band gaps increase 0.66 eV→1.42 eV→2.7 eV→ 2.91 eV.

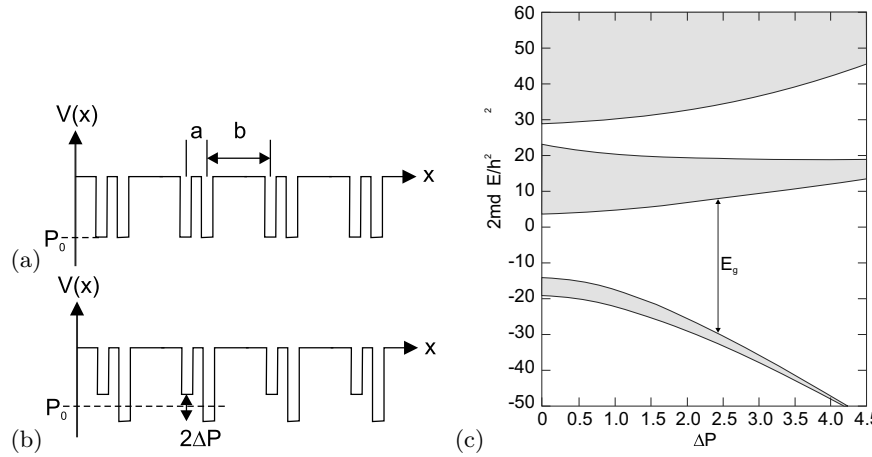

Fig. 6.22 Kronig-Penney model (along $\langle 111 \rangle$, $b/a = 3$) for a **a** IV–IV semiconductor and **b** for a III–V (or II–VI) semiconductor, **c** resulting band structure ($P_0 = -3$). d denotes the lattice constant ($d = b + a$). Adapted from [481]

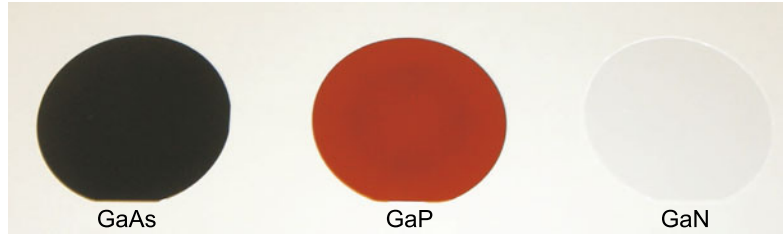

Fig. 6.23 Optical image of two inch wafers of GaAs (*left*), GaP (*center*) and ZnO (*right*). A GaN wafer would look like the ZnO wafer

This behavior can be understood within the framework of a modified Kronig-Penney model [481] (Appendix F). Double potential wells ($b/a = 3$) are chosen to mimic the diatomic planes along the $\langle 111 \rangle$ direction in the zincblende structure (Fig. 6.22a). The first investigation of such diatomic one-dimensional bandstructure was reported in [482]. Symmetric wells (depth P_0) are chosen to model covalent semiconductors and asymmetric wells with depths $P_0 \pm \Delta P$ to model partially ionic semiconductors. Results are shown in Fig. 6.22a for $P_0 = -3$. With increasing asymmetry, i.e. increasing ionicity, the band gap increases, mostly due to a downward shift of the valence band. The case of III–V (II–VI) semiconductors is reached for $\Delta P \approx 2$ (4). The calculation of effective masses in [481] is incorrect and has been rectified in [483]; the effective mass increases monotonically with ΔP.

In Fig. 6.23, the visual impression of 2" wafers of GaAs, GaP and GaN on white paper is shown. GaAs (and GaSb) is opaque since the band gap is below the visible spectral range. GaP has a band gap in the green and appears red, GaN has a band gap in the ultra-violet and thus appears transparent. As can be seen from Table 6.3, the anion sequence Sb, As, P, and N leads to smaller lattice constant and higher ionicity. A notable deviation from this rule is InN whose band gap (0.7 eV) is much smaller than that of InP [484].

6.5 Alloy Semiconductors

In alloy semiconductors [166], the size of the band gap and the character of the band gap will depend on the composition. The dependence of the band gap on the ternary composition is mostly nonlinear and can usually be expressed with a bowing parameter b that is mostly positive. For a compound $A_xB_{1-x}C$

Table 6.3 Comparison of band gap, lattice constant and ionicity of gallium–group V semiconductors for various anions. Lattice constant for GaN has been recalculated for a cubic cell

Anion	E_g (eV)	a_0 (nm)	f_i
N	3.4	0.45	0.50
P	2.26	0.545	0.33
As	1.42	0.565	0.31
Sb	0.72	0.61	0.26

the band gap is written as

$$E_g(A_xB_{1-x}C) = E_g(BC) + x\left[E_g(AC) - E_g(BC)\right] - b\,x\,(1-x)\ . \tag{6.30}$$

Even on the virtual crystal approximation (VCA) level (Sect. 3.7.3) a nonzero bowing parameter b is predicted. However, a more thorough analysis shows that the bowing cannot be treated adequately within VCA and is due to the combined effects of volume deformation of the band structure with the alloy lattice constant, charge exchange in the alloy with respect to the binary end components, a structural contribution due to the relaxation of the cation–anion bond lengths in the alloy and a small contribution due to disorder [485]. The discussion of Sect. 6.12.3 is related.

The Si_xGe_{1-x} alloy has diamond structure for all concentrations and the position of the conduction-band minimum in $\mathbf{k}$-space switches from L to X at about $x = 0.15$ (Fig. 6.24a). However, for all concentrations the band structure is indirect. The $In_xGa_{1-x}As$ alloy has zincblende structure for all compositions. The band gap is direct and decreases with a bowing parameter of $b = 0.6\,\text{eV}$ [486] (Fig. 6.24b). This means that for $x = 0.5$ the band gap is 0.15 eV smaller than expected from a linear interpolation between GaAs and InAs, as reported by various authors [487].

If one binary end component has a direct band structure and the other is indirect, a transition occurs from direct to indirect at a certain composition. An example is $Al_xGa_{1-x}As$ where GaAs is direct and AlAs is indirect. For all concentrations the crystal has zincblende structure. In Fig. 6.24c, the Γ, L and X conduction-band minima for ternary $Al_xGa_{1-x}As$ are shown. Up to an aluminum concentration of $x = 0.4$ the band structure is direct. Above this value the band structure is indirect with the conduction-band minimum being at the X-point. The particularity of $Al_xGa_{1-x}As$ is that its lattice constant is almost independent of x. For other alloys lattice match to GaAs or InP substrates is only obtained for specific compositions, as shown in Fig. 6.25. The band gap bowing in the group-III–nitride system has been discussed in [488].

If the two binary end components have different crystal structure, a phase transition occurs at a certain composition (range). An example is $Mg_xZn_{1-x}O$, where ZnO has wurtzite structure and MgO has rocksalt structure. The band gap is shown in Fig. 6.24d. In this case, each phase has its own bowing parameter.

All alloys of Fig. 6.24b–d have mixed cations. The band gap also varies upon anion substitution in a similar way as shown in Fig. 6.26 for ternary alloys with the cation Zn and the chalcogenides S, Se, Te and O.

6.6 Amorphous Semiconductors

Since the crystal lattice in an amorphous semiconductor is not periodic, the concept of $\mathbf{k}$-space and the related concepts such as band structure $E(\mathbf{k})$ break down at least partially. The density of states, however, remains a meaningful and useful quantity (Sect. 6.13.2).

In a perfectly crystalline semiconductor the eigenenergies of the states in the bands are real. An amorphous semiconductor can be modeled using a spectrum of complex energies [495]. In Fig. 6.27 the band structure of crystalline silicon is shown next to that calculated for amorphous silicon with $\alpha = 0.05$.

6.7 Temperature Dependence of the Band Gap

The band gap of a semiconductor typically decreases with increasing temperature. A direct visual impression can be obtained from the *same* LED chain at room temperature and dipped into liquid nitrogen (Fig. 6.28). Experimental data of band gap versus temperature are shown in Fig. 6.29 for bulk Si and ZnO.

The reasons for the temperature variation of the band gap are the change of electron–phonon interaction and the expansion of the lattice. The temperature coefficient may be written as

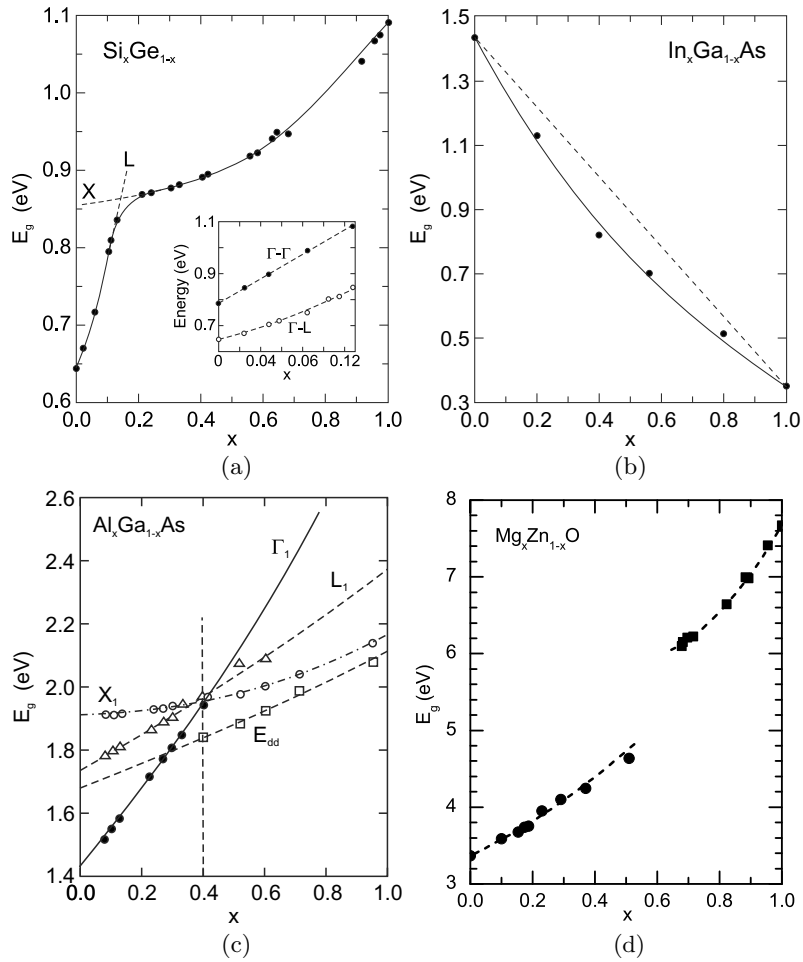

Fig. 6.24 **a** Band gap of Si_xGe_{1-x} alloy ($T = 296\,K$) with a change from the conduction-band minimum at L (Ge-rich) to X. The *inset* depicts the transition energy of the indirect (Γ–L) and direct (Γ–Γ) absorption edge for low Si content. Adapted from [489]. **b** Band gap (at room temperature) of $In_xGa_{1-x}As$. The *solid line* is an interpolation with bowing ($b = 0.6\,eV$) and the *dashed line* is the linear interpolation. Data from [486]. **c** Band gap (at room temperature) in the ternary system $Al_xGa_{1-x}As$. For $x < 0.4$ the alloy is a direct, for $x > 0.4$ an indirect, semiconductor. E_{dd} denotes the energy position of a deep donor (cf. Sect. 7.7.6). Adapted from [490]. **d** Band gap (at room temperature) in the ternary system $Mg_xZn_{1-x}O$. Data (from spectroscopic ellipsometry [491, 492]) are for hexagonal wurtzite phase (*circles*), and Mg-rich cubic rocksalt phase (*squares*). *Dashed lines* are fits to data with a different bowing parameter for each phase

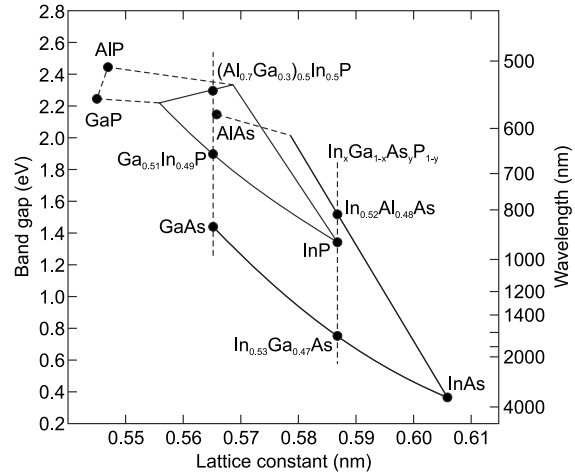

Fig. 6.25 Band gap versus lattice constant for $Ga_xIn_{1-x}P$ and $Al_xIn_{1-x}P$ (lattice matched to GaAs) as well as for $In_xAl_{1-x}As$ and $In_xGa_{1-x}As$ alloys (lattice matched to InP)

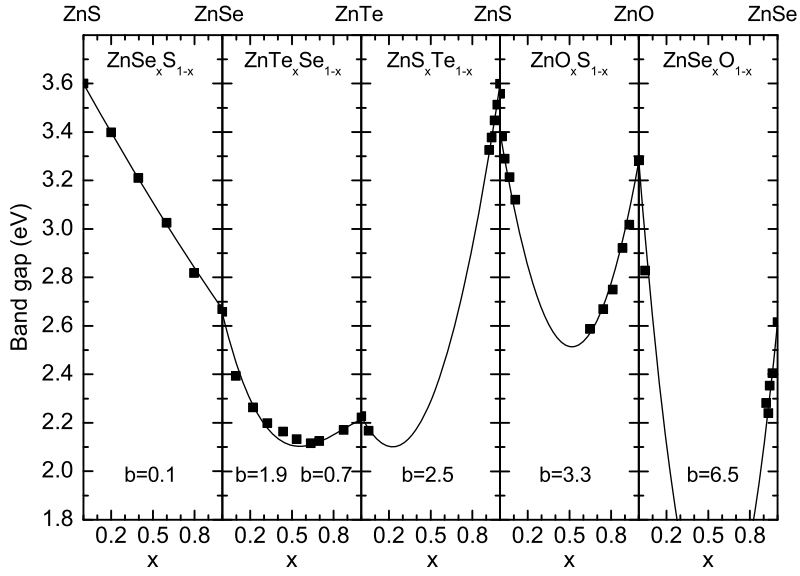

Fig. 6.26 Band gap of various Zn-based alloys with mixing in the anion sublattice. The *lines* are fits with (6.30), the bowing parameter b is labeled. Data for Zn(S,Se,Te) from [493], for Zn(O,Se/Te) from [494]

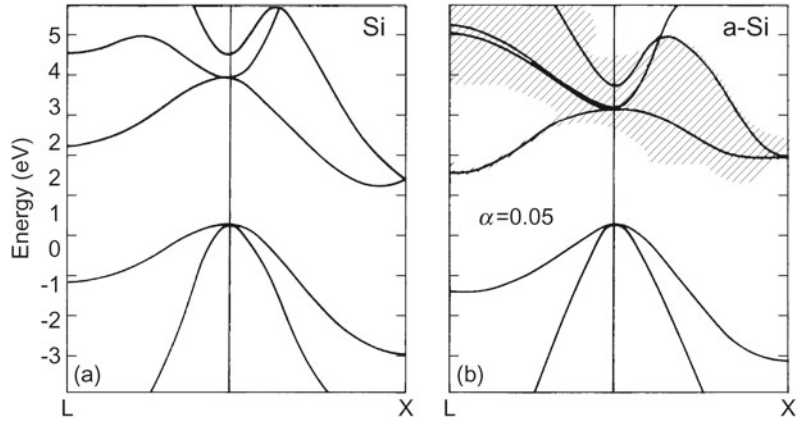

Fig. 6.27 **a** Calculated band structure of crystalline silicon. **b** Calculated band structure of amorphous silicon with $\alpha = 0.05$ (cf. (3.7)). The *solid lines* denote the real part of the energy, the *shaded areas* denote the regions with a width of twice the imaginary part of the energies centered around the real part. Adapted from [496]

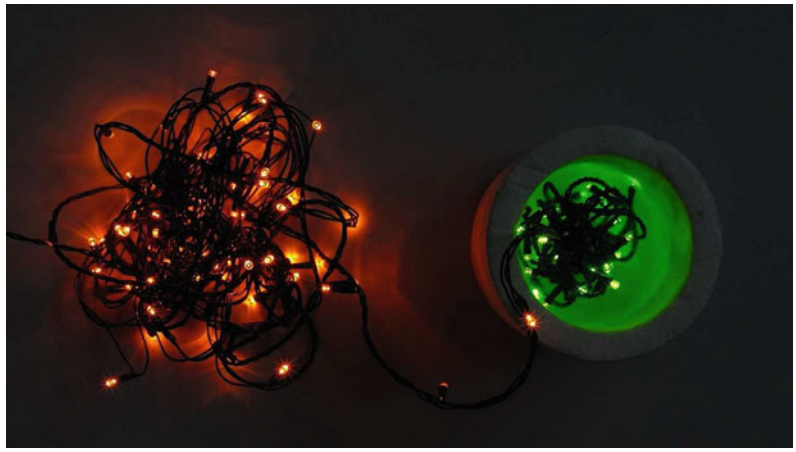

Fig. 6.28 LED chain with part at room temperature (left) and a part in a dewar filled with liquid nitrogen (right)

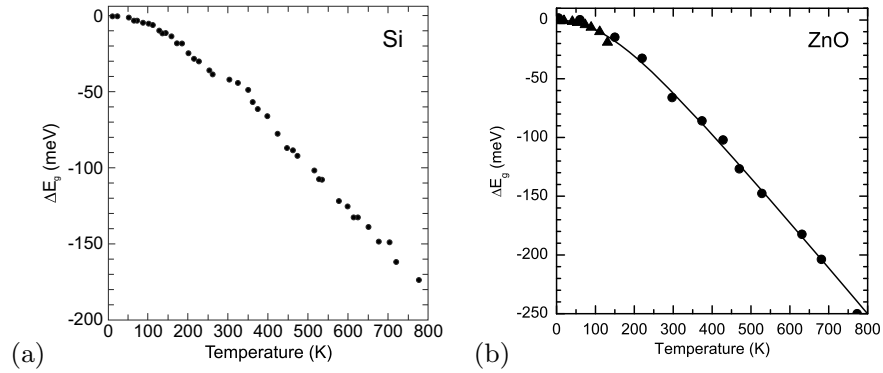

Fig. 6.29 Temperature dependence of the band gap of **a** Si (data from [498]) and **b** ZnO (experimental data from photoluminescence (*triangles*) and ellipsometry (*circles*)). The *solid lines* are fits with (6.34) and the parameters given in Table 6.4

$$\left(\frac{\partial E_g}{\partial T}\right)_p = \left(\frac{\partial E_g}{\partial T}\right)_V - \frac{\alpha}{\beta}\left(\frac{\partial E_g}{\partial p}\right)_T , \tag{6.31}$$

where α is the volume coefficient of thermal expansion and β is the volume compressibility. A recommendable discussion of the thermodynamic role of the band gap as chemical potential for the mass action law (7.12), entropy contributions and its temperature dependence can be found in [497].

An anomaly is present for the lead salts (PbS, PbSe, PbTe) for which the temperature coefficient is positive (Fig. 6.30a). Theoretical calculations [499] show that both terms in (6.31) are positive for the lead salts. The L_6^+ and L_6^- levels (see Fig. 6.12) shift as a function of temperature in such a way that their separation increases (Fig. 6.30b).

Also in copper and silver halides [500, 501] (Fig. 6.31a) and chalcopyrites [502] (Fig. 6.31b) the increase of band gap with increasing temperature has been found, sometimes only for a certain temperature range. This effect is attributed to the p-d electron hybridization in the valence band with Cu 3d electrons and to even stronger effect with Ag 4d electrons.

For many semiconductors the temperature dependence can be described with the empirical, three-parameter Varshni formula [503],

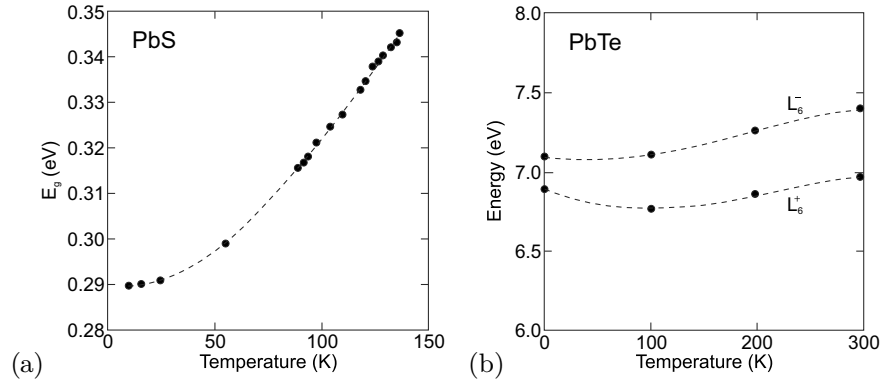

Fig. 6.30 **a** Band gap versus temperature for PbS. **b** Theoretical position of L_6^+ and L_6^- as a function of temperature for PbTe. Adapted from [468]

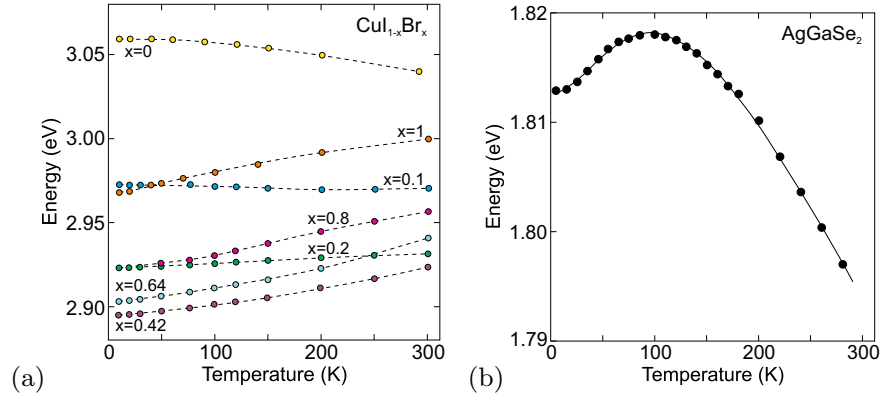

Fig. 6.31 **a** Band gap versus temperature for zincblende $CuI_{1-x}Br_x$ alloys with various compositions x (including binary CuI and CuBr) as labeled. *Dashed lines* are guide to the eye. Adapted from [500]. **b** Band gap vs. temperature for chalcopyrite $AgGaSe_2$. *Solid line* is fit with two-oscillator Bose-Einstein model. Adapted from [502]

$$E_g(T) = E_g(0) - \frac{\alpha\, T^2}{T + \beta}\,, \tag{6.32}$$

where $E_g(0)$ is the band gap at zero temperature. A more precise and physically motivated formula (based on a Bose-Einstein phonon model [504]) has been given in [505]

$$E_g(T) = E_g(0) - \frac{\alpha_B\, \Theta_B}{2}\left[\coth\left(\frac{\Theta_B}{2T}\right) - 1\right] = E_g(0) - \frac{\alpha_B\, \Theta_B}{\exp(\Theta_B/T) - 1}\,, \tag{6.33}$$

where α_B is a coupling constant and $k\Theta_B$ is a typical phonon energy; typical values are given in Table 6.4. This model reaches a better description of the fairly flat dependence at low temperatures.

The more elaborate model of [506] takes into account a more variable phonon dispersion, including optical phonons, and proposes the four-parameter formula

$$E_g(T) = E_g(0) - \alpha\,\Theta\left[\frac{1 - 3\,\Delta^2}{\exp\,(2/\gamma) - 1} + \frac{3\,\Delta^2}{2}\left(\sqrt[6]{1+\beta} - 1\right)\right] \tag{6.34}$$

Table 6.4 Parameters for the temperature dependence of the band gap according to 6.33 (Si, GaAs: [505], GaN: [507], ZnO: [508]) and (6.34) for various semiconductors

	α (10^{-4} eV/K)	Θ (K)	Δ	α_B (10^{-4} eV/K)	Θ_B (K)
Si	3.23	446	0.51	2.56	296
Ge	4.13	253	0.49		
GaAs	4.77	252	0.43	5.16	310
GaN	6.14	586	0.40	4.05	370
InP	3.96	274	0.48		
InAs	2.82	147	0.68		
ZnSe	5.00	218	0.36		
ZnO	3.8	659	0.54	5.9	616

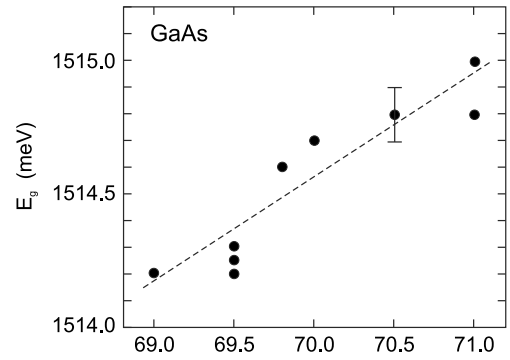

Fig. 6.32 Band gap of GaAs (at $T = 10\,\mathrm{K}$) as a function of the Ga isotope content. *Dashed line* is linear fit. Adapted from [509]

$$\beta = \frac{\pi^2}{3\,(1+\Delta^2)}\,\gamma^2 + \frac{3\Delta^2-1}{4}\,\gamma^3 + \frac{8}{3}\,\gamma^4 + \gamma^6$$
$$\gamma = 2\,T/\Theta\,,$$

where α is the high-temperature limiting magnitude of the slope (of the order of several 10^{-4} eV/K), Θ is an effective average phonon temperature and Δ is related to the phonon dispersion. Δ takes typically values between zero (Bose-Einstein model) and 3/4 [506].

6.8 Isotope Dependence of the Band Gap

The band edge slightly depends on the isotope composition of semiconductor, as shown for GaAs in Fig. 6.32. The effect is discussed in detail in [509].

6.9 Electron Dispersion

6.9.1 Equation of Electron Motion

The equation of motion for the electron in the band structure is no longer given by Netwon's law $\mathbf{F} = \mathrm{d}(m\mathbf{v})/\mathrm{d}t$ as in vacuum. Instead, the propagation of quantum-mechanical electron wave packets has to be considered. Their group velocity is given by ($v_g = \partial\omega/\partial k$)

$$\mathbf{v} = \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k}) \ , \tag{6.35}$$

where $\nabla_{\mathbf{k}}$ is the gradient with respect to $\mathbf{k}$. Through the dispersion relation the influence of the crystal and its periodic potential on the motion enters the equation.

An electric field $\mathcal{E}$ acts on an electron during the time δt the work $\delta E = -e\mathcal{E} v_g \, \delta t$. This change in energy is related to a change in k via $\delta E = \mathrm{d}E/\mathrm{d}k \, \delta k = \hbar v_g \, \delta k$. Thus, we arrive at $\hbar \, \mathrm{d}k/\mathrm{d}t = -e\mathcal{E}$. For an external force we thus have

$$\hbar \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = -e \, \mathbf{E} = \mathbf{F} \ . \tag{6.36}$$

Thus, the crystal momentum $\mathbf{p} = \hbar\mathbf{k}$ takes the role of the momentum. A more rigorous derivation can be found in [451].

In the presence of a magnetic field $\mathbf{B}$ the equation of motion is

$$\hbar \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = -e \, \mathbf{v} \times \mathbf{B} = -\frac{e}{\hbar} (\nabla_{\mathbf{k}} E) \times \mathbf{B} \ . \tag{6.37}$$

The motion in a magnetic field is thus perpendicular to the gradient of the energy, i.e. the energy of the electron does not change. It oscillates therefore on a surface of constant energy perpendicular to $\mathbf{B}$.

6.9.2 *Effective Mass of Electrons*

From the free-electron dispersion $E = \hbar^2 k^2/(2m)$ the mass of the particle is inversely proportional to the curvature of the dispersion relation, i.e. $m = \hbar^2/(\mathrm{d}^2 E/\mathrm{d}k^2)$. This relation will now be generalized for arbitrary dispersion relations. The (inverse) tensor of the effective mass is defined as

$$(m^{*-1})_{ij} = \frac{1}{\hbar^2} \frac{\partial^2 E}{\partial k_i \, \partial k_j} \ . \tag{6.38}$$

The equation $\mathbf{F} = m^* \, \dot{\mathbf{v}}$ must be understood as a tensor equation, i.e. for the components of the force $F_i = m^*_{ij} \, a_j$. Force and acceleration must no longer be collinear. In order to find the acceleration from the force, the inverse of the effective-mass tensor must be used, $\mathbf{a} = (\mathbf{m}^*)^{-1} \, \mathbf{F}$.

In Fig. 6.33 the energy dispersion of the (lowest) conduction band in a typical semiconductor, the related electron velocity and the effective mass are shown schematically.

In (6.22) the ratio of the effective mass and the free-electron mass is of the order of $m^*/m \approx U/\lambda$, the ratio of the free particle energy and the band gap. For typical semiconductors, the width of the (valence) band is of the order of 20 eV, and the gap is about 0.2–2 eV. Thus, the effective mass is expected to be 10–100 times smaller than the free-electron mass. Additionally, the relation $m^* \propto E_\mathrm{g}$ is roughly fulfilled (Fig. 6.34).

From so-called $\mathbf{k} \cdot \mathbf{p}$ theory [510] (see Appendix H) the effective electron mass is predicted to be related to the momentum matrix element $\mathbf{p}_\mathrm{cv}$

$$\mathbf{p}_\mathrm{cv} = \langle c|\mathbf{p}|v\rangle = \int_{\Omega_0} u^*_{\mathrm{c},\mathbf{k}}(\mathbf{r}) \, \mathbf{p} \, u_{\mathrm{c},\mathbf{k}}(\mathbf{r}) \, \mathrm{d}^3\mathbf{r} \ , \tag{6.39}$$

with Ω_0 being the unit cell volume and the Bloch functions $|c\rangle$ and $|v\rangle$ of the conduction and valence band, respectively, given as

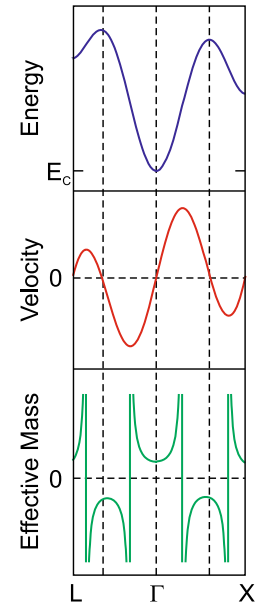

Fig. 6.33 Schematic diagram of the electron dispersion $E(\mathbf{k})$ in a typical semiconductor (*blue*) and corresponding carrier velocity ($\propto \partial E/\partial k$) (*red*) and effective mass ($\propto 1/(\partial^2 E/\partial k^2)$) (*green*)

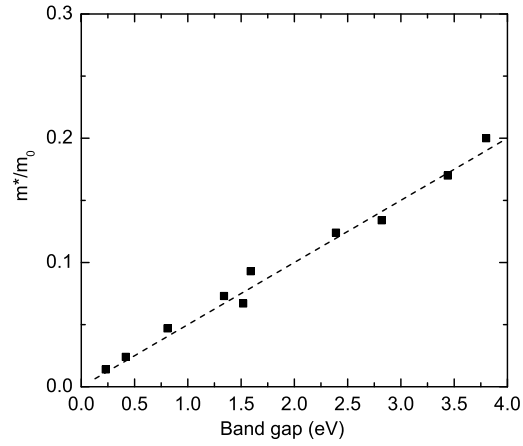

Fig. 6.34 Effective electron mass (in units of the free-electron mass m_0) as a function of the (low-temperature) band gap for several (direct band gap) semiconductors. The *dashed line* fulfills $m^*/m_0 = E_g/20\,\mathrm{eV}$

$$|c\rangle = u_{c,\mathbf{k}_c}(\mathbf{r})\ \exp(\mathrm{i}\,\mathbf{k}_c\mathbf{r}) \tag{6.40a}$$

$$|v\rangle = u_{v,\mathbf{k}_v}(\mathbf{r})\ \exp(\mathrm{i}\,\mathbf{k}_v\mathbf{r})\ . \tag{6.40b}$$

Typically, the $\mathbf{k}$-dependence of the matrix element is small and neglected. The momentum matrix element will also be important for optical transitions between the valence and conduction bands (Sect. 9.6). Other related quantities that are often used are the energy parameter E_P

$$E_P = \frac{2\,|\mathbf{p}_{cv}|^2}{m_0}\ , \tag{6.41}$$

and the bulk momentum matrix element M_b^2 that is given by

$$M_b^2 = \frac{1}{3}\,|\mathbf{p}_{cv}|^2 = \frac{m_0}{6}\,E_P\ . \tag{6.42}$$

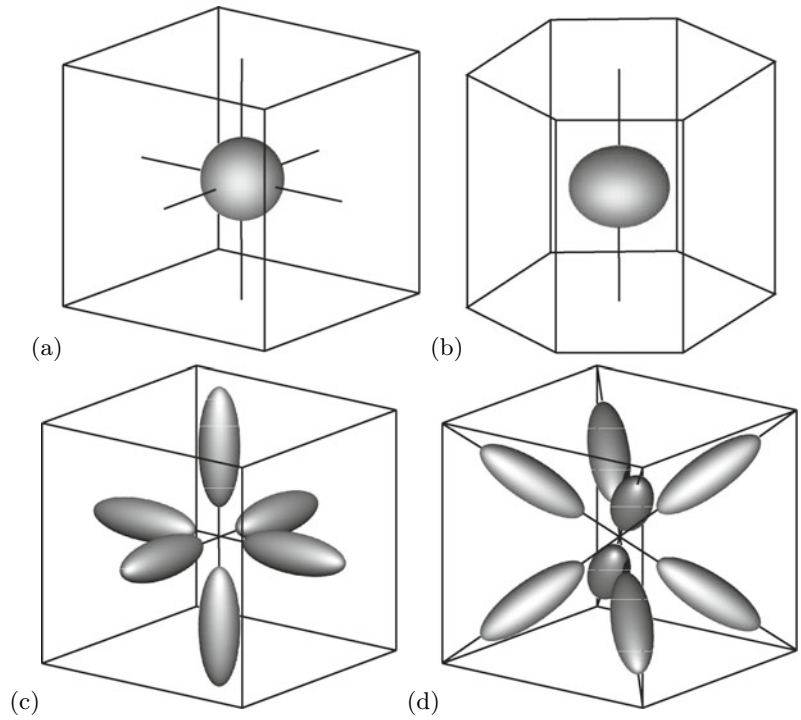

Fig. 6.35 Energy isosurfaces in **k**-space in the vicinity of the conduction-band minima for **a** GaAs with isotropic (spehrical) minimum at Γ-point, **b** ZnO with anisotropic (ellipsoidal) minimum at Γ-point (anisotropy exaggerated), **c** silicon with six equivalent anisotropic minima ($m_l/m_t = 5$ not to scale) along ⟨100⟩ and **d** germanium with eight equivalent anisotropic minima along ⟨111⟩. The cube indicates the ⟨100⟩ directions for the cubic materials. For the wurtzite material (part **b**) the vertical direction is along [00.1]

The electron mass is given by[4]

$$\frac{m_0}{m_e^*} = 1 + \frac{E_P}{3}\left(\frac{2}{E_g} + \frac{1}{E_g + \Delta_0}\right) = 1 + E_P \frac{E_g + 2\Delta_0/3}{E_g\left(E_g + \Delta_0\right)} \approx 1 + \frac{E_P}{E_g + \Delta_0/3} \approx \frac{E_P}{E_g} . \tag{6.43}$$

Comparison with the fit from Fig. 6.34 yields that E_P is similar for all semiconductors [511] and of the order of 20 eV (InAs: 22.2 eV, GaAs: 25.7 eV, InP: 20.4 eV, ZnSe: 23 eV, CdS: 21 eV).

In silicon there are six equivalent conduction-band minima. The surfaces of equal energy are schematically shown in Fig. 6.35c. The ellipsoids are extended along the ⟨100⟩ direction because the longitudinal mass (along the Δ path) is larger than the transverse mass in the two perpendicular directions (Table 6.5). For example, the dispersion relation in the vicinity of one of the minima is given as (k_x^0 denotes the position of one of the conduction-band minima close to a X-point)

$$E(\mathbf{k}) = \hbar^2 \left(\frac{(k_x - k_x^0)^2}{2\, m_l} + \frac{k_y^2 + k_z^2}{2\, m_t} \right) . \tag{6.44}$$

For germanium surfaces of constant energy around the eight conduction-band minima in the ⟨111⟩ directions are depicted in Fig. 6.35d. The longitudinal and the transverse masses are again different. For GaAs, the conduction-band dispersion around the Γ point is isotropic, thus the surface of constant energy is simply a sphere (Fig. 6.35a). In wurtzite semiconductors the conduction-band minimum is at the Γ-point. The mass along the c-axis is typically smaller than the mass within the (00.1) plane [512]

[4] Δ_0 is the spin-orbit splitting discussed in Sect. 6.10.2.

Table 6.5 Longitudinal direction of effective mass ellipsoid, longitudinal and transverse effective electron mass in several semiconductors. For the density of states mass $m_{d,e}$ see (6.72). Mass values in units of the free electron mass m_0

	Long. dir.	m_l	m_t	m_l/m_t	$m_{d,e}$	Ref.
C	⟨100⟩	1.4	0.36	3.9	1.9	[514]
Si	⟨100⟩	0.98	0.19	5.16	1.08	[515]
Ge	⟨111⟩	1.59	0.082	19.4	0.88	[515]
ZnO	[00.1]	0.21	0.25	0.88		[516]
CdS	[00.1]	0.15	0.17	0.9		[517]

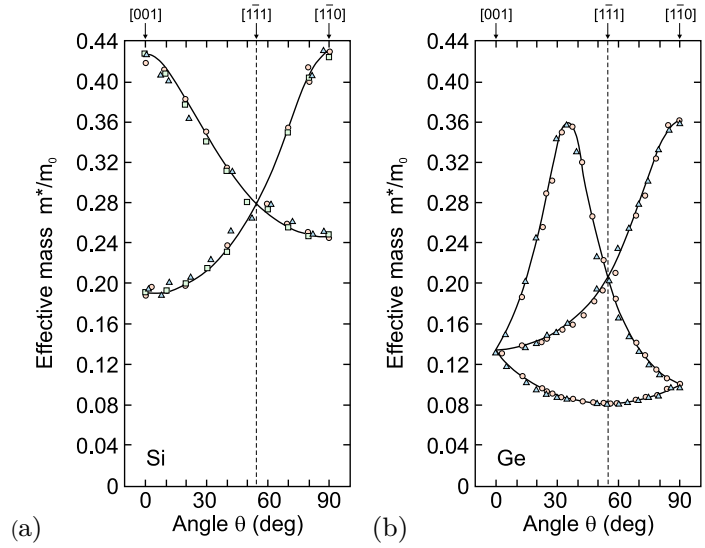

Fig. 6.36 Effective electron mass from cyclotron resonance experiments (at $T = 4\,\mathrm{K}$) on **a** Si and **b** Ge for the magnetic field in the (110) plane and various azimuthal directions θ. Experimental data (*symbols*) and fits (*solid lines*) using (6.45) with **a** $m_l = 0.98$, $m_t = 0.19$ and **b** $m_l = 1.58$, $m_t = 0.082$. Adapted from [515]

($m_l/m_t \approx 0.8$ for ZnO [513]), see Fig. 6.35b. In [512] also an anisotropy within the (00.1) plane is predicted.

The directional dependence of the mass can be measured with cyclotron resonance experiments with varying direction of the magnetic field. In Fig. 6.36, the field **B** is in the (110) plane with different azimuthal directions. When the (static) magnetic field makes an angle ϑ with the longitudinal axis of the energy surface, the effective mass is given as [518]

$$\frac{1}{m^*} = \sqrt{\frac{\cos^2\vartheta}{m_t^2} + \frac{\sin^2\vartheta}{m_t\, m_l}}\ . \tag{6.45}$$

6.9.3 Nonparabolicity of Electron Mass

The dispersion around the conduction-band minimum is only parabolic for small **k**. The further away the wavevector is from the extremum, the more the actual dispersion deviates from the ideal parabola (see, e.g., Fig. 6.10). This effect is termed *nonparabolicity*. Typically, the energy increases less quickly with k than in the parabolic model. This can be described in a so-called two-level model with the dispersion relation

$$\frac{\hbar^2 k^2}{2m_0^*} = E\left(1 + \frac{E}{E_0^*}\right) , \tag{6.46}$$

where $E_0^* > 0$ parameterizes the amount of nonparabolicity (a parabolic band corresponds to $E_0^* = \infty$). The nonparabolic dispersion for GaAs is shown in Fig. 6.37a. The curvature is reduced for larger $\mathbf{k}$ and thus the effective mass is energy dependent and increases with the energy. Equation (6.46) leads to the energy-dependent effective mass

$$m^*(E) = m_0^*\left(1 + \frac{2E}{E_0^*}\right) , \tag{6.47}$$

where m_0^* denotes here the effective mass at $\mathbf{k} = 0$. Theory and experimental data for the effective electron mass of GaAs are shown in Fig. 6.37b.

6.10 Holes

6.10.1 Hole Concept

Holes are missing electrons in an otherwise filled band. A Schrödinger-type wave-equation for holes (unoccupied electron states) was derived by Heisenberg [70] to interpret Hall effect data. The hole concept is useful to describe the properties of charge carriers at the top of the valence band. The hole is a new quasi-particle whose dispersion relation is schematically shown in Fig. 6.38 in relation to the dispersion of electrons in the valence band.

The wavevector of the hole (filled circle in Fig. 6.38) is related to that of the 'missing' electron (empty circle in Fig. 6.38) by $k_\mathrm{h} = -k_\mathrm{e}$. The energy is $E_\mathrm{h}(k_\mathrm{h}) = -E_\mathrm{e}(k_\mathrm{e})$, assuming that $E_\mathrm{V} = 0$, otherwise $E_\mathrm{h}(k_\mathrm{h}) = -E_\mathrm{e}(k_\mathrm{e}) + 2E_\mathrm{V}$. The hole energy is larger for holes that are further away from the top of the valence band, i.e. the lower the energy state of the missing electron. The velocity of the hole, $v_\mathrm{h} = \hbar^{-1}\,\mathrm{d}E_\mathrm{h}/\mathrm{d}k_\mathrm{h}$, is the same, $v_\mathrm{h} = v_\mathrm{e}$, and the charge is positive, $+e$. The effective mass of the

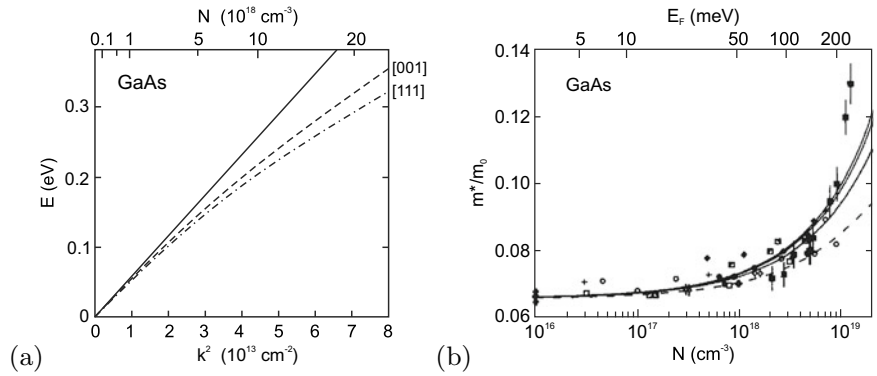

Fig. 6.37 **a** Dispersion relations for the conduction band of GaAs. The *solid line* is parabolic dispersion (constant effective mass). The *dashed* (*dash-dotted*) *line* denotes the dispersion for $\mathbf{k}$ along [001] ([111]) from a five-level $\mathbf{k}\cdot\mathbf{p}$ model (5LM). **b** Cyclotron resonance effective mass of electrons in GaAs as a function of the Fermi level (upper abscissa) and the corresponding electron concentration (lower abscissa). The *dashed line* is from a 2LM according to (6.47) with $E_0^* = 1.52\,\mathrm{eV}$. The *solid lines* are for a 5LM for the three principal directions of the magnetic field. The *symbols* represent experimental data from different sources. Data from [519]

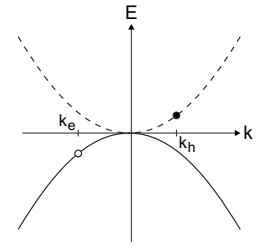

Fig. 6.38 Hole dispersion *dashed line* in relation to the electron dispersion in the valence band (*solid line*)

hole is positive at the top of the valence band, $m_h^* = -m_e^*$. Therefore, the drift velocities of an electron and hole are opposite to each other. The resulting current, however, is the same.

6.10.2 Hole Dispersion Relation

The valence band at the Γ-point is 3-fold degenerate. The band developed from the atomic (bonding) p states; the coupling of the spin $s = 1/2$ electrons with the orbital angular momentum $l = 1$ leads to a total angular momentum $j = 1/2$ and $j = 3/2$. The latter states are degenerate at Γ in zincblende bulk material and are called *heavy* holes (hh) for $m_j = \pm 3/2$ and *light* holes (lh) for $m_j = \pm 1/2$ due to their different dispersion (Fig. 6.39a). The two ($m_j = \pm/2$) states of the $j = 1/2$ state are split-off from these states by an energy Δ_0 due to spin-orbit interaction and are called *split-off* (s-o) holes. The spin-orbit interaction increases with increasing atomic order number Z of the anion since the electrons are located preferentially there (Fig. 6.40). A detailed discussion of the spin-orbit splitting in zincblende semiconductors is given in [520].

All three holes have different mass. In the vicinity of the Γ-point the dispersion for heavy and light holes can be described with (+:hh, −:lh)

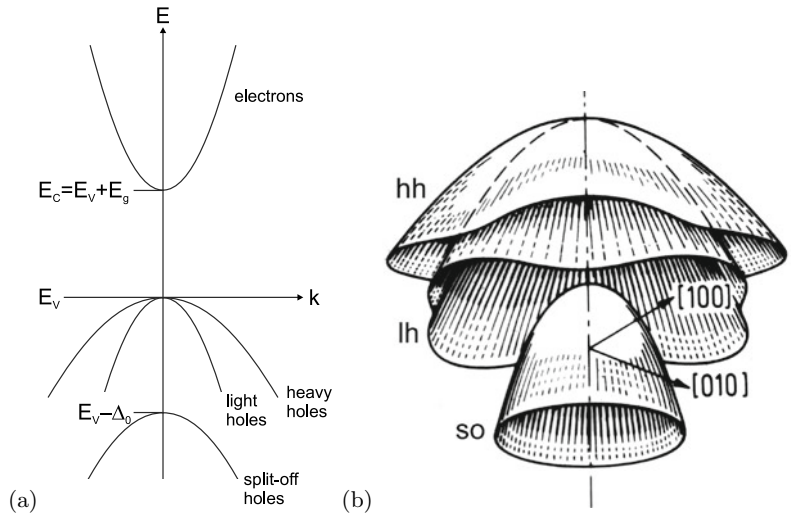

Fig. 6.39 **a** Simplified band structure with conduction band and three valence bands and (**b**) three-dimensional visualization (E versus (k_x, k_y)) of the valence bands of Ge (including warping). Part **b** from [521]

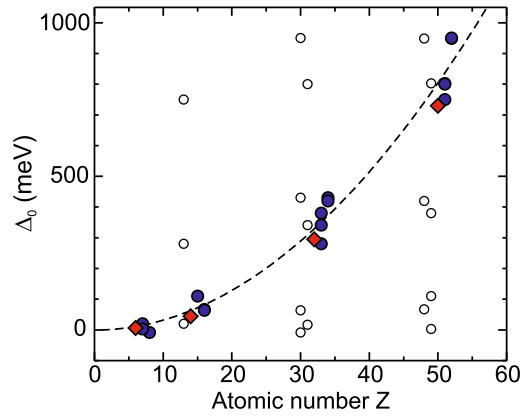

Fig. 6.40 Spin-orbit splitting Δ_0 for elemental (*diamonds*) and various III–V and II–VI (*circles*) semiconductors. The data are plotted as *empty* (*filled*) *circles* as a function of the cation (anion) order number. Obviously, Δ_0 correlates with the anion Z. The *dashed line* is proportional to Z^2

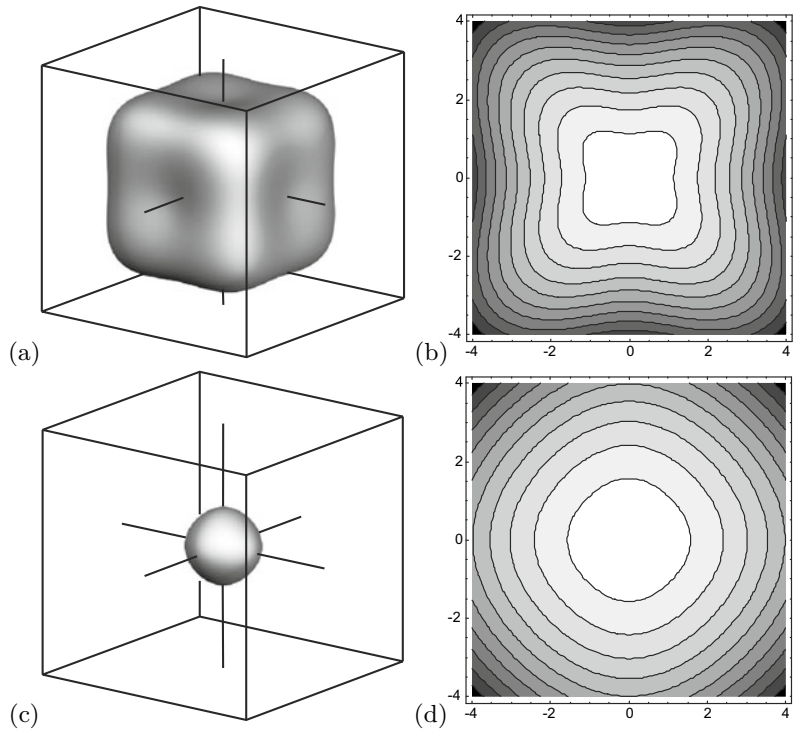

Fig. 6.41 Dispersion at the valence band edge of GaAs for **a**, **b** heavy holes and **c**,**d** light holes. **a**,**c** Constant energy surfaces and **b**,**d** isolines in the (k_x, k_y)-plane (**b** and **d** have different energy scales)

$$E(\mathbf{k}) = A\,k^2 \pm \sqrt{B^2\,k^4 + C^2\,\left(k_x^2\,k_y^2 + k_y^2\,k_z^2 + k_x^2\,k_z^2\right)}\,. \tag{6.48}$$

For heavy and light holes there is a dependence of the dispersion, i.e. the mass, in the (001) plane. This effect, sketched in Fig. 6.39b, is called *warping*. The warping at the GaAs valence-band edge is shown in Fig. 6.41. Equation (6.48) can also be expressed in terms of angular coordinates [522].

The s-o holes have the dispersion

$$E(\mathbf{k}) = -\Delta_0 + A\,k^2\,. \tag{6.49}$$

Table 6.6 Valence-band parameters (for (6.48)) A and B in units of $(\hbar^2/2m_0)$, C^2 in units of $(\hbar^2/2m_0)^2$, and Δ_0 in eV. From [164, 523, 524]

Material	A	B	C^2	Δ_0
C	−4.24	−1.64	9.5	0.006
Si	−4.28	−0.68	24	0.044
Ge	−13.38	−8.5	173	0.295
GaAs	−6.9	−4.4	43	0.341
InP	−5.15	−1.9	21	0.11
InAs	−20.4	−16.6	167	0.38
ZnSe	−2.75	−1.0	7.5	0.43

Values for A, B, C^2 and Δ_0 for a number of semiconductors are given in Table 6.6. The valence-band structure is often described with the Luttinger parameters γ_1, γ_2, and γ_3 that can be represented through A, B, and C via

$$\frac{\hbar^2}{2m_0}\gamma_1 = -A \tag{6.50a}$$

$$\frac{\hbar^2}{2m_0}\gamma_2 = -\frac{B}{2} \tag{6.50b}$$

$$\frac{\hbar^2}{2m_0}\gamma_3 = \frac{\sqrt{B^2 + C^2/3}}{2}\,. \tag{6.50c}$$

The mass of holes in various directions can be derived from (6.48). The mass along the [001] direction, i.e. $\hbar^2/(\partial^2 E(\mathbf{k})/\partial k_x^2)$ for $k_y = 0$ and $k_z = 0$, is

$$\frac{1}{m_{\mathrm{hh}}^{100}} = \frac{2}{\hbar^2}(A + B) \tag{6.51a}$$

$$\frac{1}{m_{\mathrm{lh}}^{100}} = \frac{2}{\hbar^2}(A - B)\,. \tag{6.51b}$$

The anisotropy of hole masses has been investigated with cyclotron resonance experiments (Fig. 6.42). For θ being the angle between the magnetic field and the [001] direction, the effective heavy hole (upper sign) and light hole (lower sign) mass in cubic semiconductors is given by [515]

$$m^* = \frac{\hbar^2}{2}\,\frac{1}{A \pm \sqrt{B^2 + C^2/4}} \times \left\{ \frac{C^2\,(1 - 3\cos^2\theta)^2}{64\sqrt{B^2 + C^2/4}\left[A \pm \sqrt{B^2 + C^2/4}\right]} + \ldots \right\}\,. \tag{6.52}$$

For $C^2 = 0$ the hole bands are isotropic, as is obvious from (6.48). In this case $\gamma_2 = \gamma_3$, the so-called spherical approximation. The average of the hole masses over all directions is

$$\frac{1}{m_{\mathrm{hh}}^{\mathrm{av}}} = \frac{2}{\hbar^2}\left(A + B\left[1 + \frac{2\,C^2}{15\,B^2}\right]\right) \tag{6.53a}$$

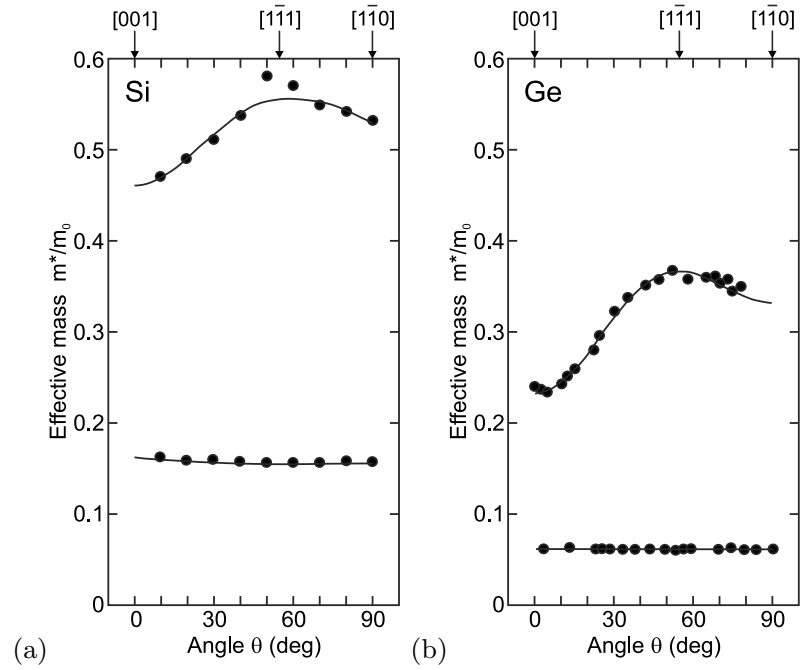

Fig. 6.42 Effective hole masses from cyclotron resonance experiments ($T = 4\,\mathrm{K}$) for heavy and light holes in **a** Si and **b** Ge for the magnetic field in the (110) plane and various azimuthal directions θ. Experimental data (*symbols*) and fits (*solid lines*) using (6.52). Adapted from [515]

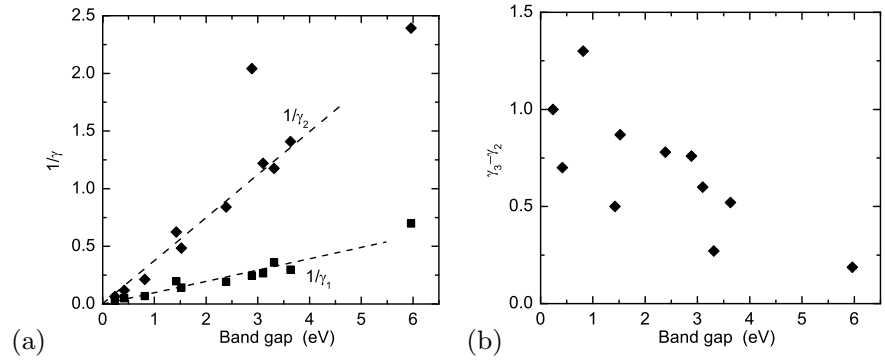

Fig. 6.43 Luttinger parameters for various III-V semiconductors versus their band gap. **a** Inverse values of γ_1 (*squares*) and γ_2 (*diamonds*). *Dashed lines* are guides to the eye. **b** $\gamma_3 - \gamma_2$ versus band gap

$$\frac{1}{m_{\mathrm{lh}}^{\mathrm{av}}} = \frac{2}{\hbar^2}\left(A - B\left[1 + \frac{2\,C^2}{15\,B^2}\right]\right) . \tag{6.53b}$$

Similar to the correlation of the electron mass with the band gap (Fig. 6.34), the Luttinger parameters are correlated with the band gap as shown in Fig. 6.43. The parameters $1/\gamma_1$ and $1/\gamma_2$ increase about linearly with E_g. The parameter $\gamma_3 - \gamma_2$, which is responsible for the valence band warping, decreases with increasing band gap.

6.10.3 Valence-Band Fine Structure

In Fig. 6.44, the schematic structure of the band edges for zincblende structure semiconductors is shown. The s-o holes in the zincblende structure are split-off due to the spin-orbit interaction Δ_{so}, the Γ_8 band is degenerate (heavy and light holes). Degeneracies for the holes are removed in the wurtzite and chalcopyrite structures by the additional crystal field splitting Δ_{cf} due to the anisotropy between the a- and c-axes. Typically, e.g. for CdS, the topmost valence band in the wurtzite structure has Γ_9 symmetry (allowed optical transitions only for $\mathbf{E} \perp \mathbf{c}$); an exception is ZnO for which the two upper bands are believed to be reversed. In the chalcopyrite structure optical transitions involving the Γ_6 band are only allowed for $\mathbf{E} \perp \mathbf{c}$. The three hole bands are usually labeled A, B, and C from the top of the valence band.

The energy positions of the three bands (with respect to the position of the Γ_{15} band) in the presence of spin-orbit interaction and crystal field splitting are given within the quasi-cubic approximation [525] by

$$E_1 = \frac{\Delta_{so} + \Delta_{cf}}{2} \tag{6.54a}$$

$$E_{2,3} = \pm\sqrt{\left(\frac{\Delta_{so} + \Delta_{cf}}{2}\right)^2 - \frac{2}{3}\,\Delta_{so}\,\Delta_{cf}}\,. \tag{6.54b}$$

In chalcopyrites the crystal field splitting is typically negative (Fig. 6.45). It is approximately linearly related to $1 - \eta$ (for $\eta = c/2a$ see Sect. 3.4.6).

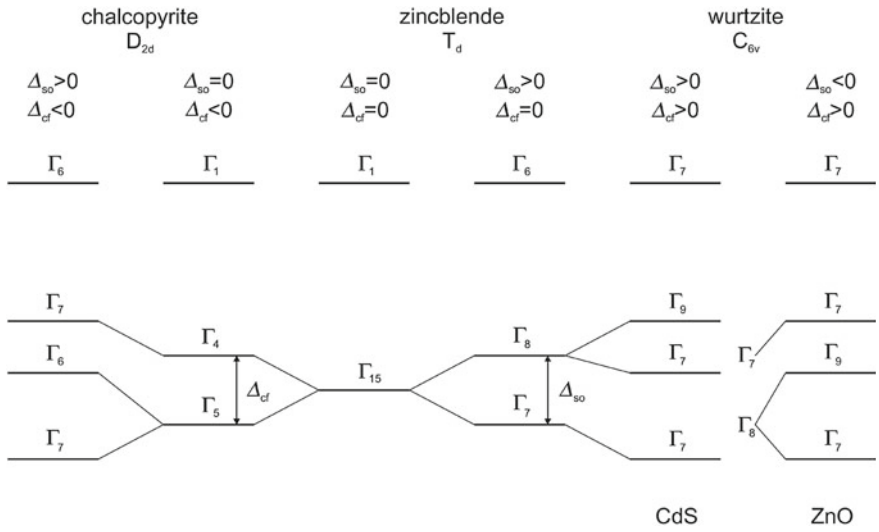

Fig. 6.44 Schematic band structure of zincblende and the valence-band splitting due to spin-orbit interaction Δ_{so} and crystal field splitting Δ_{cf} for chalcopyrites (typically $\Delta_{cf} < 0$, see Fig. 6.45) and wurtzites. For the wurtzites the situation is schematically shown for CdS ($\Delta_{so} = 67$ meV, $\Delta_{cf} = 27$ meV) (or GaN) and ZnO ($\Delta_{so} = -8.7$ meV, $\Delta_{cf} = 41$ meV)

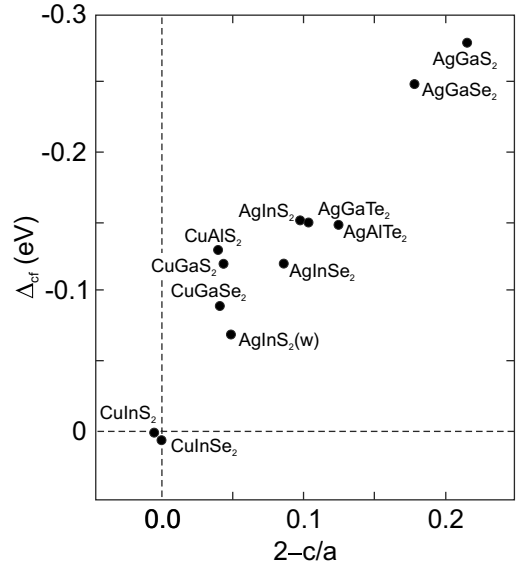

Fig. 6.45 Crystal field splitting Δ_{cf} for various chalcopyrite compounds versus the tetragonal distortion $2 - c/a = 2\,(1 - \eta)$. *Dash-dotted line* represents $\Delta_{cf} = 1.5\,b\,(2 - c/a)$ for $b = 1$ eV. Data from [526]

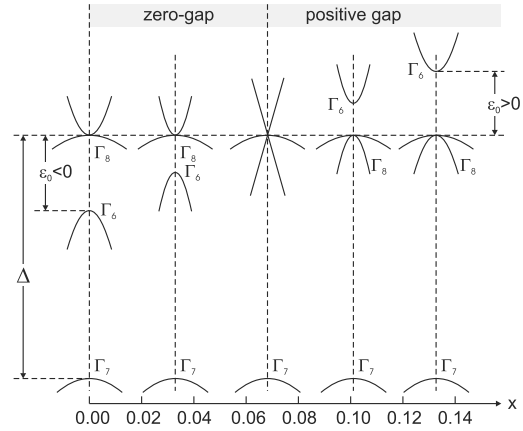

Fig. 6.46 Schematic band structure of zincblende with vanishing energy gap for the ternary compounds $Mn_xHg_{1-x}Te$. Note the linear dispersion for the zero-gap case at $x \approx 0.07$

6.11 Band Inversion

In certain compounds, typically mixing a semiconductor with a semimetal [527, 528], the band gap can shrink to zero (zero-gap semiconductor) and even become negative in the sense that the s-type Γ_6 symmetry (conduction) band is inverted below the Γ_8 (p-type) valence-band edge. HgTe is a classical example for such material as shown in Fig. 6.46, but similar effects are also present in other semiconductors, for example various chalcopyrites [529]. Remember that such band structures are topologically non-trivial (cf. Sect. 6.2.6).

For the zero-gap case, the dispersion of the two crossing bands is linear (like for graphene, cf. Sect. 13.1.2). The dielectric function of zero-gap semiconductors is discussed in [530].

For the $Cd_xHg_{1-x}Te$ system, around the zero-gap concentration of $x \approx 0.16$, the change from normal to inverted band structure will occur also as a function of temperature [531] as shown in Fig. 6.47. Such effect had been described already 50 years ago for (Pb,Sn)Te at the L-point (cf. Sect. 6.3.6) in [532].

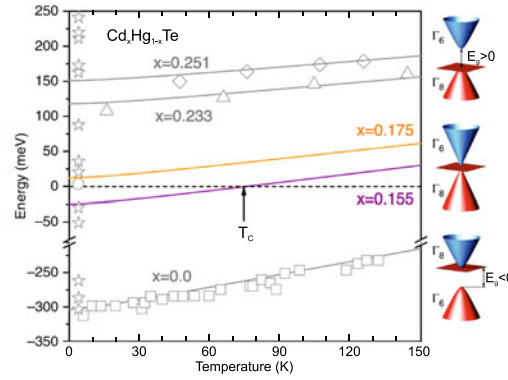

Fig. 6.47 Band gap of $Cd_xHg_{1-x}Te$ for various alloy compositions and temperatures. On the right, the schematic band structure of (Hg,Cd)Te with positive, zero and negative band gap is shown. Adapted from [531], reprinted under a Creative Commons Attribution (CC BY 4.0) licence

6.12 Strain Effects on the Band Structure

A mechanical strain (or equivalently stress) causes changes in the bond lengths. Accordingly, the band structure is affected. These effects have been exhaustively treated in [533, 534]. For small strain, typically $\epsilon \lesssim 0.01$ the shift of the band edges is linear with the strain, for large strain it becomes nonlinear [535]. Often homogeneous strain is assumed, the effect of inhomogeneous strain is discussed in [536].

6.12.1 Strain Effect on Band Edges

In a direct-gap zincblende material the position of the conduction-band edge is only affected by the hydrostatic component of the strain

$$E_C = E_C^0 + a_c \left(\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}\right) = E_C^0 + a_c \,\mathrm{Tr}(\epsilon) \,, \tag{6.55}$$

where $a_c < 0$ is the conduction-band hydrostatic deformation potential and E_C^0 is the conduction-band edge of the unstrained material. Similarly, the valence-band edge is

$$E_V = E_V^0 + a_v \,\mathrm{Tr}(\epsilon) \,, \tag{6.56}$$

where $a_v > 0$ is the valence-band hydrostatic deformation potential. Therefore the band gap increases by

$$\Delta E_g = a \,\mathrm{Tr}(\epsilon) = a \left(\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}\right) \,, \tag{6.57}$$

with $a = a_C - a_V$. Such linear behavior upon hydrostatic pressure has been found for many semiconductors and is shown in Fig. 6.48a for $Ga_{0.92}In_{0.08}As$. The anomaly for N-doping is discussed below in Sect. 6.12.3. In Fig. 6.49 the dependence of the direct and indirect gaps of GaAs is shown. The dependence of the direct gap on pressure is non-linear, that on the density is linear [537].

Biaxial and shear strains affect the valence bands and lead to shifts and splitting of the heavy and light holes at the Γ-point:

$$E_{v,hh/lh} = E_v^0 \pm E_{\epsilon\epsilon} \tag{6.58a}$$

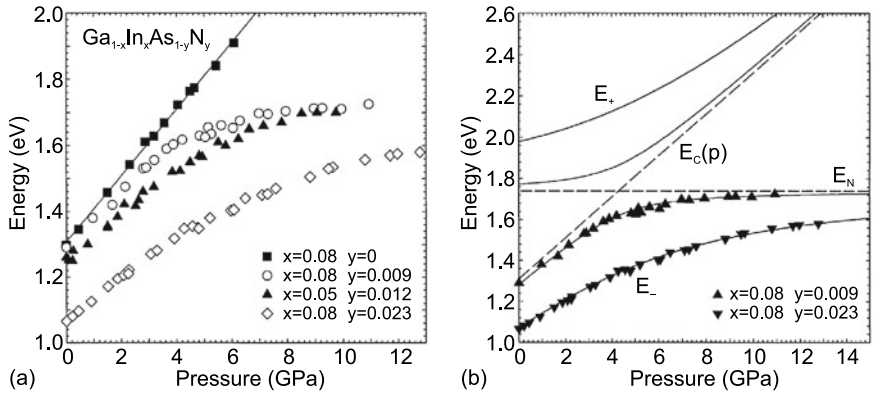

Fig. 6.48 **a** Dependence of the band gap of $Ga_{0.92}In_{0.08}As$ alloy (*squares*) and nitrogen-doped (Ga,In)As on (compressive) hydrostatic pressure, determined by photomodulated transmission at $T = 295\,\mathrm{K}$. **b** Pressure dependence of band gap for two (Ga,In)(As,N) samples together with model calculation (6.62). The coupling parameter is $V = 0.12\,\mathrm{eV}$ (0.4 eV) for a nitrogen content of 0.9% (2.3%). Adapted from [538]

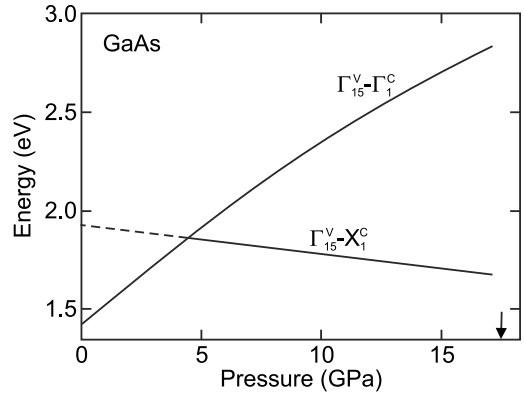

Fig. 6.49 Dependence of the direct Γ_{15}^{V}–Γ_{1}^{C} and indirect Γ_{15}^{V}–X_{1}^{C} band gap of GaAs ($T = 300\,\mathrm{K}$) on pressure. *Solid lines* are interpolations of experimental data, *dashed line* is extrapolation to $p = 0$. The crossing of the direct and indirect band gap occurs at 4.2 GPa. The *arrow* denotes the pressure of the phase transition from zincblende to an orthorhomic structure around 17 GPa. Adapted from [537]

$$E_{\epsilon\epsilon}^{2} = b^{2}/2\left[\left(\epsilon_{xx}-\epsilon_{yy}\right)^{2}+\left(\epsilon_{yy}-\epsilon_{zz}\right)^{2}+\left(\epsilon_{xx}-\epsilon_{zz}\right)^{2}\right] + d^{2}\left[\epsilon_{xy}^{2}+\epsilon_{yz}^{2}+\epsilon_{xz}^{2}\right],$$

where E_{v}^{0} denotes the bulk valence-band edge. b and d are the optical deformation potentials. For compressive strain the heavy-hole band is above the light-hole band. For tensile strain there is strong mixing of the bands (Fig. 6.50). In Table 6.7 the deformation potentials for some III–V semiconductors are listed. Typical values are in the eV regime.

In a wurtzite crystal, seven (or eight) deformation potentials are needed that are termed a (for the change of band gap with hydrostatic strain, again $a = a_C - a_V$) and D_1–D_6 (for the valence band structure) [539, 540].

In Si and Ge, three deformation potentials, termed a, b, d, are needed for the valence band and two for each conduction band minimum, Ξ_u and Ξ_d [541]. The energy position of the i-th conduction-band edge (with unit vector $\mathbf{a}_i$ pointing to the valley) is

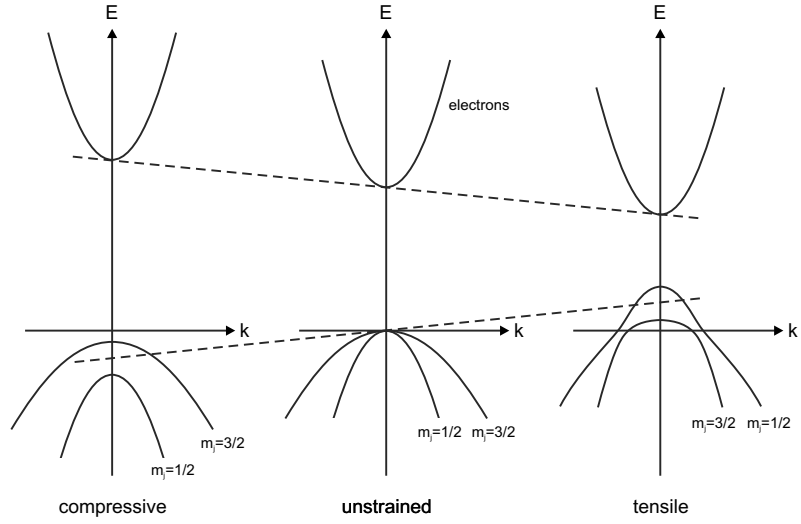

Fig. 6.50 Schematic band structure of GaAs in unstrained state (*center*) and under compressive and tensile biaxial strain as labeled. *Dashed lines* indicate shift of band edges due to hydrostatic part of strain

Table 6.7 Deformation potentials for some III–V semiconductors. All values in eV

Material	a	b	d
GaAs	−9.8	−1.7	−4.6
InAs	−6.0	−1.8	−3.6

Table 6.8 Deformation potentials for silicon and germanium. All values in eV from [542]

material	$\Xi_{\mathrm{d}}^{(\Delta)}$	$\Xi_{\mathrm{u}}^{(\Delta)}$	$\Xi_{\mathrm{d}}^{(\mathrm{L})}$	$\Xi_{\mathrm{u}}^{(\mathrm{L})}$	a	b	d
Si	1.1	10.5	−7.0	18.0	2.1	−2.33	−4.75
Ge	4.5	9.75	−4.43	16.8	2.0	−2.16	−6.06

$$E_{\mathrm{C},i} = E_{\mathrm{C},i}^{0} + \Xi_{\mathrm{d}}\,\mathrm{Tr}(\boldsymbol{\epsilon}) + \Xi_{\mathrm{u}}\,\mathbf{a}_i\,\boldsymbol{\epsilon}\,\mathbf{a}_i\ , \tag{6.59}$$

where $E_{\mathrm{C},i}^{0}$ denotes the energy of the unstrained conduction-band edge. The deformation potentials for Si and Ge are given in Table 6.8.

6.12.2 Strain Effect on Effective Masses

In the presence of strain the band edges are shifted (cf. Sect. 6.12). Since the electron mass is related to the band gap, it is expected that the mass will also be effected. In the presence of hydrostatic strain ϵ_{H} the electron mass is [543] (cf. to (6.43) for $\epsilon_{\mathrm{H}} \to 0$)

$$\frac{m_0}{m_{\mathrm{e}}^*} = 1 + \frac{E_{\mathrm{P}}}{E_{\mathrm{g}} + \Delta_0/3}\left[1 - \epsilon_{\mathrm{H}}\left(2 + \frac{3a}{E_{\mathrm{g}} + \Delta_0/3}\right)\right]\ , \tag{6.60}$$

with a being the hydrostatic deformation potential and $\epsilon_{\mathrm{H}} = \mathrm{Tr}(\epsilon)$. In [543], formulas are also given for biaxial and shear strain and also for hole masses. Since the effective mass enters the mobility, the electrical conductivity depends on the stress state of the semiconductor (piezoresistivity, see Sect. 8.3.14).

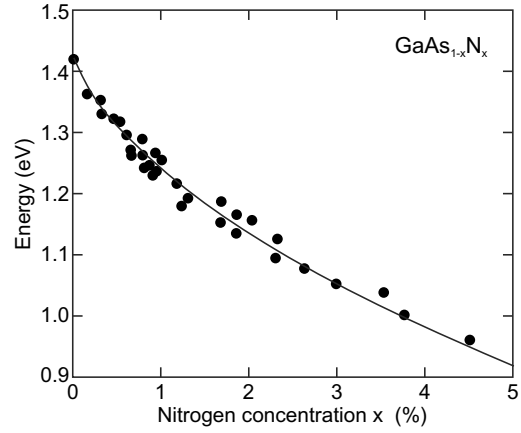

Fig. 6.51 Bandgap of $GaAs_{1-x}N_x$, experimental data from various sources (*symbols*) and model (*curve*) according to (6.62) with $V = V_0\sqrt{x}$ for $V_0 = 2.7\,\text{eV}$. Adapted from [545]

6.12.3 Interaction With a Localized Level

The normal dependence of the band gap on hydrostatic pressure is linear and given by (6.57). (Ga,In)As containing nitrogen exhibits a remarkable deviation from this behavior as shown in Fig. 6.48a. This is due to the interaction of the continuum states of the conduction band with the electron level of the isoelectronic nitrogen impurity (Sect. 7.7.9) E_N, being within the conduction band. For GaAs it is 0.2 eV above the conduction band edge E_C. This phenomenon has been investigated theoretically within microscopic detail [544]. Within a simple 'band anticrossing' two-level model, the coupling of the pressure-dependent conduction band edge E_C and the nitrogen level can be obtained from the Eigenwert equation

$$\begin{vmatrix} E - E_C & V \\ V & E - E_N \end{vmatrix} = 0\,, \tag{6.61}$$

V being the coupling constant. The determinant vanishes for

$$E_\pm = \frac{1}{2}\left(E_C + E_N \pm \sqrt{(E_C - E_N)^2 + 4V^2}\right)\,. \tag{6.62}$$

Here the weak pressure dependence of E_N is neglected for simplicity. This model can explain the pressure dependence of the band gap of (Ga,In)As:N fairly well [538] (Fig. 6.48b). The coupling parameter V is in the order of a few 0.1 eV for small nitrogen content. In photomodulated reflection also the E_+ levels can be observed [545]. The anti-crossing model can also model the dependence of the bandgap of $GaAs_{1-x}N_x$ on the nitrogen concentration [545] (Fig. 6.51).

6.13 Density of States

6.13.1 General Band Structure

The dispersion relation yields how the energy of a (quasi-) particle depends on the **k** vector. Now we want to know how many states are at a given energy. This quantity is called the *density of states* (DOS) and is written as $D(E)$. It is defined in an infinitesimal sense such that the number of states between

E and $E + \delta E$ is $D(E)\delta E$. In the vicinity of the extrema of the band structure many states are at the same energy such that the density of states is high.

The dispersion relation of a band will be given as $E = E(\mathbf{k})$. If several bands overlap, the densities of state of all bands need to be summed up. The density of states at the energy $\tilde{E}$ for the given band is

$$D(\tilde{E})\,\mathrm{d}E = 2\int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi/L)^3}\,\delta(\tilde{E} - E(\mathbf{k}))\,, \tag{6.63}$$

where, according to (5.5), $(2\pi/L)^3$ is the $\mathbf{k}$-space volume for one state. The factor 2 is for spin degeneracy. The integral runs over the entire $\mathbf{k}$-space and selects only those states that are at $\tilde{E}$. The volume integral can be converted to a surface integral over the isoenergy surface $S(\tilde{E})$ with $E(\mathbf{k}) = \tilde{E}$. The volume element $\mathrm{d}^3\mathbf{k}$ is written as $\mathrm{d}^2S\,\mathrm{d}\mathbf{k}_\perp$. The vector $\mathrm{d}\mathbf{k}_\perp$ is perpendicular to $S(\tilde{E})$ and proportional to $\nabla_\mathbf{k}E(\mathbf{k})$, i.e. $\mathrm{d}E = |\nabla_\mathbf{k}E(\mathbf{k})|\,\mathrm{d}\mathbf{k}_\perp$.

$$D(\tilde{E}) = 2\int_{S(\tilde{E})} \frac{\mathrm{d}^2S}{(2\pi/L)^3}\,\frac{1}{|\nabla_\mathbf{k}E(\mathbf{k})|}\,. \tag{6.64}$$

In this equation, the dispersion relation is explicitly contained. At band extrema the gradient diverges, however, in three dimensions the singularities are integrable and the density of states takes a finite value. The corresponding peak is named a van-Hove singularity. The concept of the density of states is valid for all possible dispersion relations, e.g. for electrons, phonons or photons.

The density of states for the silicon band structure (see Fig. 6.9a) is shown in Fig. 6.52.

6.13.2 Amorphous Semiconductors

If disorder is introduced, the density of states is modified as shown in Fig. 6.53 for amorphous germanium using a calculation with complex eigenenergies. The defects, as compared to the perfect lattice, introduced states in the band gap and generally wash out the sharp features from the crystalline DOS.

Several models exist for the defect level distributions within the band gap. The first model was the Mott model which has band tails at the valence and conduction band edges [547]. In the Cohen-Fritzsche-Ovshinsky (CFO) model [548], the band tails are more severe and overlap; the Fermi energy lies at the minimum of the density of states. In the Davis–Mott model [549] deep states were added in the gap and eventually the Marshall-Owen model [550] assumes band tails and donor- and acceptor-like deep states. The four models are schematically shown in Fig. 6.54. These model densities of states allow also the interpretation of carrier transport in amorphous semiconductors, taking into account localized and delocalized states (see Sect. 8.9).

The density of states for an amorphous semiconductor is best calculated from atomistic models, possibly averaging over many configurations. The typical features, compared to the clear band gap of a similar ordered material, are band tails due to disorder (cmp. Sect. 5.2.9) and deep levels within the gap due to specific atomic arrangements not present in ordered bulk. The most investigated system is amorphous silicon; in Fig. 6.55 a numerical calculation of the density of states is shown together with charge distribution of four states at selected energies [551]. The further the states are in the band tail, the stronger their localization is. The two most right states shown in Fig. 6.55 are not conducting.

As another example, simulations of $ZnSnO_3$ are shown in Fig. 6.56. The band tail between 0 and 0.5 eV is due to disorder of oxygen 2p orbitals [552]. At 0.9 eV a level due to under-coordinated oxygen appears. Deep levels are due to metal-metal bonds. Band tails due to chemically disordered oxygen have been experimentally observed for amorphous GIZO [553].

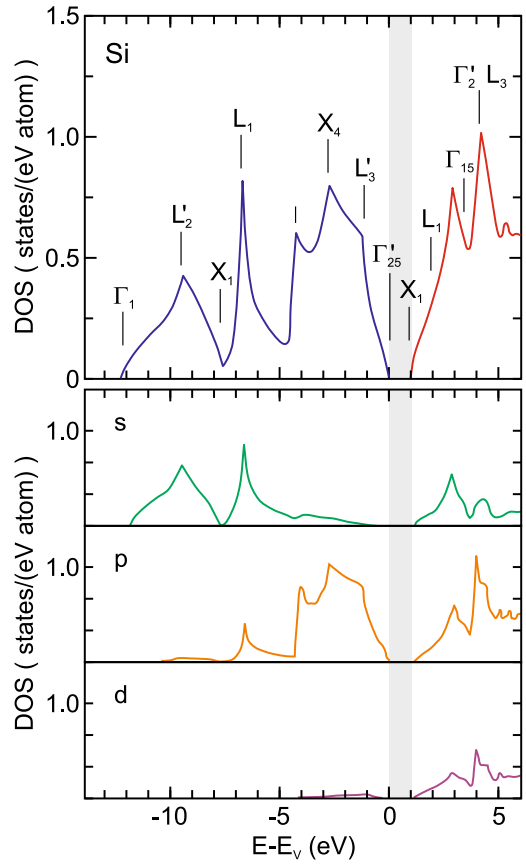

Fig. 6.52 Density of states in the silicon valence- (*blue*) and conduction-band (*red*) as obtained from theoretical calculation using empirical pseudopotentials. *Grey* regions denotes the band gap. Critical points (cf. Fig. 6.9a) are labeled. In the lower three graphs, the DOS is decomposed into contributions from different angular momentum states (s (*green*), p (*orange*) and d (*purple*)). *Top part* adapted from [546], *bottom part* adapted from [175]

6.13.3 *Free-Electron Gas*

In M dimensions, the energy states of a free-electron gas are given as

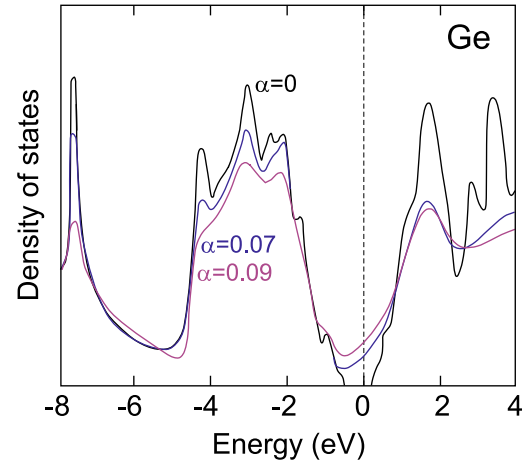

Fig. 6.53 Theoretical calculation for the density of states of amorphous Ge models as obtained for various degrees of disorder α (3.7). $\alpha = 0.09$ corresponds to a mean short-range order distance of about 2.4 lattice constants (cmp. Fig. 3.14b). Adapted from [204]

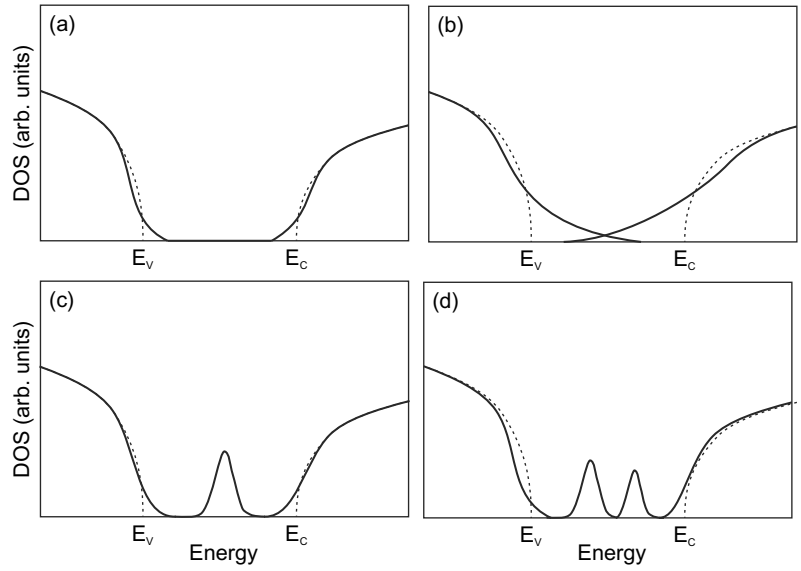

Fig. 6.54 Model density of states in amorphous semiconductors (*solid lines*) according to Mott [547], Cohen-Fritzsche-Ovshinsky [548], Davis–Mott [549] and Marshall–Owen [550]. *Dashed lines* represent the DOS of the same material without disorder

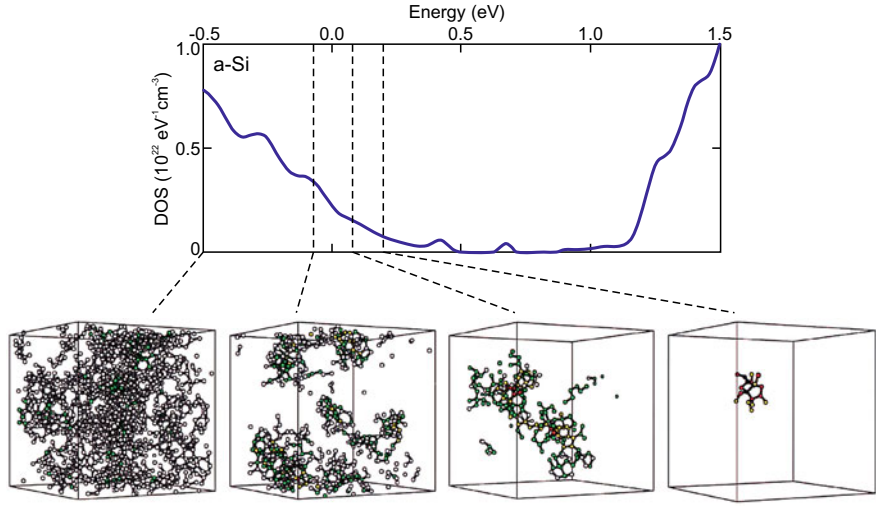

Fig. 6.55 Theoretical calculation of the density of electronic states of amorphous silicon. The charge distribution in four selected states at the indicated energies is shown, from right to left with decreasing localization. Adapted from [551]

$$E(\mathbf{k}) = \frac{\hbar^2}{2\,m^*} \sum_{i=1}^{M} k_i^2 \,. \tag{6.65}$$

The k_i can take the values $\pm\pi n/L$ (in the first Brillouin zone) with $n \leq N$, N being the number of unit cells in one dimension. These values are equidistant in $\mathbf{k}$-space. Each M-dimensional $\mathbf{k}$-point takes a volume of $(2\pi/L)^M$. The number of states $N(E_\mathrm{F})$ up to the energy $E_\mathrm{F} = \frac{\hbar^2}{2m} k_\mathrm{F}^2$ (later used as Fermi energy E_F and Fermi vector k_F) is

$$N(E_\mathrm{F}) = \frac{2}{(2\pi/L)^M} \int_{\mathbf{k}=0}^{|\mathbf{k}|=k_\mathrm{F}} \mathrm{d}^M k \,. \tag{6.66}$$

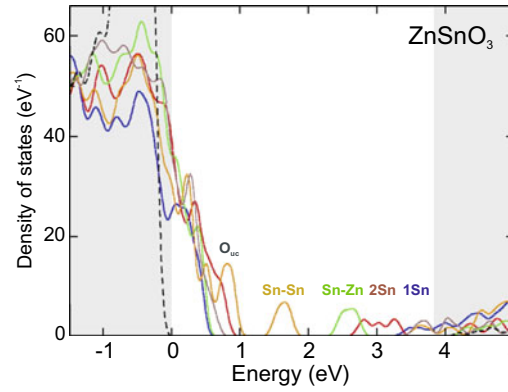

Fig. 6.56 Theoretical calculation for the density of states of crystalline (*dashed lines*, conduction and valence bands indicated by *greay areas*) and amorphous $ZnSnO_3$ with different configurations (*solid lines*). States due to under-coordinated oxygen (O_{uc}) and metal-metal bonds are labelled. Adapted from [552]

The factor 2 is for spin degeneracy, the integration runs over M dimensions. The density of states is the derivative

$$D(E) = \frac{\mathrm{d}N}{\mathrm{d}E} \,. \tag{6.67}$$

In the following, the density of states for $M = 3, 2, 1$ and zero dimensions is derived. A visualization is given in Fig. 14.1.

6.13.3.1 $M = 3$

This case relates to bulk material in which electrons are free to move in all three dimensions. Performing the integral (6.66) for $M = 3$ yields for an isotropic mass,

$$N^{3\mathrm{D}} = \frac{V}{3\pi^2}\, k_\mathrm{F}^3 = \frac{V}{3\pi^2} \left(\frac{2\, m\, E_\mathrm{F}}{\hbar^2} \right)^{3/2} \,. \tag{6.68}$$

Therefore, k_F and E_F are given by

$$k_\mathrm{F} = \left(\frac{3\pi^2\, N}{V} \right)^{1/3} \tag{6.69}$$

$$E_\mathrm{F} = \frac{\hbar^2}{2\, m^*} \left(\frac{3\pi^2\, N}{V} \right)^{2/3} \,, \tag{6.70}$$

and the density of states in three dimensions is

$$D^{3\mathrm{D}}(E) = \frac{V}{2\pi^2} \left(\frac{2\, m^*}{\hbar^2} \right)^{3/2} \sqrt{E} \,. \tag{6.71}$$

Mostly the density of states is used as density of states per volume, then the factor V in (6.71) is omitted.

If a conduction-band minimum is degenerate, a factor g_v (valley degeneracy) must be included in the density of states, i.e. $g_\mathrm{v} = 6$ for Si and $g_\mathrm{v} = 8$ for Ge ($g_\mathrm{v} = 1$ for GaAs). This factor is typically included in the mass used in (6.71) that then becomes the density of states mass $m_{\mathrm{d,e}}$. If the conduction-band minimum has cylindrical symmetry in **k**-space, such as for Si and Ge, the mass that has to be

used is

$$m_{\mathrm{d,e}} = g_{\mathrm{v}}^{2/3} \left(m_{\mathrm{t}}^2 m_{\mathrm{l}}\right)^{1/3} . \tag{6.72}$$

In the case of a degeneracy of the valence band, the states of several bands need to be summed. In bulk material, typically the heavy and light hole bands are degenerate at the Γ-point. If the split-off band is not populated because of insufficient temperature, the valence-band edge density of states is expressed by the density of states hole mass

$$m_{\mathrm{d,h}} = \left(m_{\mathrm{hh}}^{3/2} + m_{\mathrm{lh}}^{3/2}\right)^{2/3} . \tag{6.73}$$

The density of states (per volume) at the conduction and valence band edges are thus given by

$$D_{\mathrm{e}}^{\mathrm{3D}}(E) = \frac{1}{2\pi^2} \left(\frac{2\, m_{\mathrm{d,e}}}{\hbar^2}\right)^{3/2} \sqrt{E - E_{\mathrm{C}}} \, , \, E > E_{\mathrm{C}} \tag{6.74}$$

$$D_{\mathrm{h}}^{\mathrm{3D}}(E) = \frac{1}{2\pi^2} \left(\frac{2\, m_{\mathrm{d,h}}}{\hbar^2}\right)^{3/2} \sqrt{E_{\mathrm{V}} - E} \, , \, E < E_{\mathrm{V}} . \tag{6.75}$$

6.13.3.2 $M = 2$

This case is important for thin layers in which the electron motion is confined in one direction and free in a plane. Such structures are called quantum wells (see Sect. 12.3.2). We find for the 2D density of states (for each subband over which it is not summed here, including spin degeneracy)

$$N^{\mathrm{2D}} = \frac{A}{2\pi} k_{\mathrm{F}}^2 = \frac{A}{\pi} \frac{m^*}{\hbar^2} E \, , \tag{6.76}$$

where A is the area of the layer. The density of states is thus constant and given by

$$D^{\mathrm{2D}}(E) = \frac{A}{\pi} \frac{m^*}{\hbar^2} . \tag{6.77}$$

6.13.3.3 $M = 1$

The case $M = 1$ describes a quantum wire in which the electron motion is confined in two dimensions and free in only one dimension. For this case, we find for a wire of length L

$$N^{\mathrm{1D}} = \frac{2L}{\pi} k_{\mathrm{F}} = \frac{2L}{\pi} \left(\frac{2m^* E}{\hbar^2}\right)^{1/2} . \tag{6.78}$$

The density of states becomes singular at $E = 0$ and is given by (for one subband)

$$D^{\mathrm{1D}}(E) = \frac{L}{\pi} \left(\frac{2m^*}{\hbar^2}\right)^{1/2} \frac{1}{\sqrt{E}} . \tag{6.79}$$

6.13.3.4 $M = 0$

In this case electrons have no degrees of freedom, as, e.g., in a quantum dot (Sect. 14.4), and each state has a δ-like density of states at each of the quantized levels.