CAMBRIDGE

Cambridge University Press & Assessment 978-0-521-76344-8 — Semiconductor Devices for High-Speed Optoelectronics Giovanni Ghione Excerpt <u>More Information</u>

# 1 Semiconductors, alloys, heterostructures

# 1.1 Introducing semiconductors

Single-crystal semiconductors have a particularly important place in optoelectronics, since they are the starting material for high-quality sources, receivers and amplifiers. Other materials, however, can be relevant to some device classes: polycrystalline or amorphous semiconductors can be exploited in light-emitting diodes (LEDs) and solar cells; dielectrics (also amorphous) are the basis for passive devices (e.g., waveguides and optical fibers); and piezoelectric (ferroelectric) crystals such as lithium niobate are the enabling material for a class of electrooptic (EO) modulators. Moreover, polymers have been recently exploited in the development of active and passive optoelectronic devices, such as emitters, detectors, and waveguides (e.g., fibers). Nevertheless, the peculiar role of single-crystal semiconductors justifies the greater attention paid here to this material class with respect to other optoelectronic materials.

From the standpoint of electron properties, semiconductors are an intermediate step between insulators and conductors. The electronic structure of crystals generally includes a set of allowed energy bands, that electrons populate according to the rules of quantum mechanics. The two topmost energy bands are the valence and conduction band, respectively, see Fig. 1.1. At some energy above the conduction band, we find the vacuum level, i.e., the energy of an electron free to leave the crystal. In insulators, the valence band (which hosts the electrons participating to the chemical bonds) is separated from the conduction band by a large energy gap  $E_g$ , of the order of a few electronvolts (eV). Due to the large gap, an extremely small number of electrons have enough energy to be promoted to the conduction band, where they could take part into electrical conduction. In insulators, therefore, the conductivity is extremely small. In metals, on the other hand, the valence and conduction bands overlap (or the energy gap is *negative*), so that all carriers already belong to the conduction band, independent of their energy. Metals therefore have a large conductivity. In semiconductors, the energy gap is of the order of 1-2 eV, so that some electrons have enough energy to reach the conduction band, leaving *holes* in the valence band. Holes are pseudo-particles with positive charge, reacting to an external applied electric field and contributing, together with the electrons in the conduction band, to current conduction. In pure (intrinsic) semiconductors, therefore, charge transport is *bipolar* (through electrons and holes), and the conductivity is low, exponentially dependent on the gap (the larger the gap, the lower the conductivity). However, impurities can be added (dopants) to provide large numbers of electrons to

## 2 Semiconductors, alloys, heterostructures



Figure 1.1 Main features of semiconductor bandstructure.  $E_g$  is the energy gap;  $E_c$  is the conduction band edge;  $E_v$  is the valence band edge.

the conduction band (*donors*) or of holes to the valence band (*acceptors*). The resulting doped semiconductors are denoted as n-type and p-type, respectively; their conductivity can be artificially modulated by changing the amount of dopants; moreover, the dual doping option allows for the development of pn junctions, one of the basic building blocks of electronic and optoelectronic devices.

# 1.2 Semiconductor crystal structure

Crystals are regular, periodic arrangements of atoms in three dimensions. The point set  $\underline{r}$  defining the crystal nodes, corresponding to the atomic positions (Bravais lattice) satisfies the condition  $\underline{r} = k\underline{a}_1 + l\underline{a}_2 + m\underline{a}_3$ , where k, l, m are integer numbers and  $\underline{a}_1$ ,  $\underline{a}_2, \underline{a}_3$  are the *primitive vectors* denoting the *primitive cell*, see Fig. 1.2. Bravais lattices can be formed so as to fill the entire space only if the angles  $\alpha_1, \alpha_2, \alpha_3$  assume values from a discrete set (60°, 90°, 120°, or the complementary value to 360°). According to the relative magnitudes of  $a_1, a_2, a_3$  and to the angles  $\alpha_1, \alpha_2, \alpha_3, 14$  basic lattices can be shown to exist, as in Table 1.1. In semiconductors, only two lattices are technologically important at present, i.e. the *cubic* and the *hexagonal*. Most semiconductors are cubic and the hexagonal structure can be found in carbon (C), where they are the diamond and graphite crystal structures, respectively.

Three kinds of Bravais cubic lattices exist, the simple cubic (sc), the face-centered cubic (fcc) and the body-centered cubic (bcc), see Fig. 1.3. The cubic semiconductor crystal structure can be interpreted as two *shifted* and *compenetrated* fcc Bravais lattices.

Let us consider first an elementary semiconductor (e.g., Si) where all atoms are equal. The relevant cubic lattice is the *diamond lattice*, consisting of two interpenetrating

## 1.2 Semiconductor crystal structure

3

#### Table 1.1 The 14 Bravais lattices.

Name	Bravais lattices	Conditions on primitive vectors
Triclinic	1	$a_1 \neq a_2 \neq a_3, \alpha_1 \neq \alpha_2 \neq \alpha_3$
Monoclinic	2	$a_1 \neq a_2 \neq a_3, \alpha_1 = \alpha_2 = 90^\circ \neq \alpha_3$
Orthorhombic	4	$a_1 \neq a_2 \neq a_3, \alpha_1 = \alpha_2 = \alpha_3 = 90^{\circ}$
Tetragonal	2	$a_1 = a_2 \neq a_3, \alpha_1 = \alpha_2 = \alpha_3 = 90^{\circ}$
Cubic	3	$a_1 = a_2 = a_3, \alpha_1 = \alpha_2 = \alpha_3 = 90^{\circ}$
Trigonal	1	$a_1 = a_2 = a_3, \alpha_1 = \alpha_2 = \alpha_3 < 120^\circ \neq 90^\circ$
Hexagonal	1	$a_1 = a_2 \neq a_3, \alpha_1 = \alpha_2 = 90^\circ, \alpha_3 = 120^\circ$



Figure 1.2 Semiconductor crystal structure: definition of the primitive cell.



Figure 1.3 Cubic Bravais lattices: (a) simple, (b) body-centered, (c) face-centered.

fcc Bravais lattices, displaced along the body diagonal of the cubic cell by onequarter the length of the diagonal, see Fig. 1.4. Since the length of the diagonal is  $d = a |\hat{x} + \hat{y} + \hat{z}| = a\sqrt{3}$ , the displacement of the second lattice is described by the vector

$$\underline{s} = \frac{a\sqrt{3}}{4}\frac{\hat{x} + \hat{y} + \hat{z}}{\sqrt{3}} = \frac{a}{4}\left(\hat{x} + \hat{y} + \hat{z}\right).$$

# 1.2.1 The Miller index notation

The Miller indices are a useful notation to denote planes and reference directions within a lattice. The notation (h, k, l), where h, k, l are integers, denotes the set of parallel planes that intercepts the three points  $\underline{a_1}/h$ ,  $\underline{a_2}/k$  and  $\underline{a_3}/l$ , or some multiple thereof, while [h, k, l] in square brackets is the direction orthogonal to plane (h, k, l).

4

## Semiconductors, alloys, heterostructures



Figure 1.4 The diamond lattice as two cubic face-centered interpenetrating lattices. The pale and dark gray points represent the atoms falling in the basic cell.



Figure 1.5 Examples of planes and directions according to the Miller notation.

Additionally,  $\{h, k, l\}$  is a family of planes with symmetries and  $\langle h, k, l\rangle$  is the related direction set. In cubic lattices, the primitive vectors coincide with the Cartesian axes and  $a_1 = a_2 = a_3 = a$ , where *a* is the lattice constant; in this case, we simply have  $[h, k, l] \equiv h\hat{x} + k\hat{y} + l\hat{z}$  where  $\hat{x}, \hat{y}$  and  $\hat{z}$  are the Cartesian unit vectors.

To derive the Miller indices from the plane intercepts in a cubic lattice, we normalize with respect to the lattice constant (thus obtaining a set of integers (H, K, L)), take the reciprocal  $(H^{-1}, K^{-1}, L^{-1})$  and finally multiply by a minimum common multiplier so as to obtain a set (h, k, l) such as  $h : k : l = H^{-1} : K^{-1} : L^{-1}$ . Notice that a zero index corresponds to an intercept point at infinity. Examples of important planes and directions are shown in Fig. 1.5.

**Example 1.1:** Identify the Miller indices of the following planes, intersecting the coordinate axes in points (normalized to the lattice constant): (a) x = 4, y = 2, z = 1; (b) x = 10, y = 5,  $z = \infty$ ; (c) x = 3.5,  $y = \infty$ ,  $z = \infty$ ; (d) x = -4, y = -2, z = 1.

We take the reciprocal of the intercept, and then we multiply by the minimum common multiplier, so as to obtain an integer set with minimum module. In case (a), the reciprocal set is (1/4, 1/2, 1), with minimum common multiplier 4, leading to the Miller indices (1, 2, 4). In case (b), the reciprocals are (1/10, 1/5, 0) with Miller indices (1, 2, 0). In case (c), the plane is orthogonal to the *z* axis, and the Miller indices simply are (1, 0, 0). Finally, case (d) is similar to case (a) but with negative intercepts; according to the Miller notation we overline the indices rather than using a minus sign; we thus have  $(\overline{1}, \overline{2}, 4)$ .

#### 1.2 Semiconductor crystal structure

5

# 1.2.2 The diamond, zinc-blende, and wurtzite semiconductor cells

The cubic diamond cell includes 8 atoms; in fact, if we consider Fig. 1.6, the corner atoms each contribute to eight adjacent cells, so that only 8/8 = 1 atom belongs to the main cell. The atoms lying on the faces belong half to the main cell, half to the nearby ones, so that only 6/2 = 3 atoms belong to the main cell. Finally, the other (internal) 4 atoms belong entirely to the cell. Therefore, the total number of atoms in a cell is 1 + 3 + 4 = 8. In the diamond cell, each atom is connected to the neighbours through a tetrahedral bond. All atoms are the same (C, Si, Ge...) in the diamond lattice, while in the so-called *zinc-blende lattice* the atoms in the two fcc constituent lattices are different (GaAs, InP, SiC...). In particular, the corner and face atoms are metals (e.g., Ga) and the internal atoms are nonmetals (e.g., As), or vice versa.

In the diamond or zinc-blende lattices the Miller indices are conventionally defined with respect to the cubic cell of side a. Due to the symmetry of the tetrahedral atom bonds, planes (100) and (110), etc. have two bonds per side, while planes (111) have three bonds on the one side, two on the other. Moreover, the surface atom density is different, leading, for example, to different etch velocities.

Some semiconductors, such as SiC and GaN, have the hexagonal *wurtzite* crystal structure. Hexagonal lattices admit many *polytypes* according to the stacking of successive atom layers; a large number of polytypes exists, but only a few have interesting semiconductor properties (e.g. 4H and 6H for SiC). The wurtzite cell is shown in Fig. 1.7, including 12 equivalent atoms. In the ideal lattice, one has

$$|\underline{a}_3| = c, \quad |\underline{a}_1| = |\underline{a}_2| = a, \quad \frac{c}{a} = \sqrt{\frac{8}{3}} \approx 1.633.$$

Some properties of semiconductor lattices are shown in Table 1.2.<sup>1</sup> It can be noted that wurtzite-based semiconductors are often anisotropic (uniaxial) and have two dielectric constants, one parallel to the *c*-axis, the other orthogonal to it.



Figure 1.6 The diamond (left) and zinc-blende (right) lattices.

<sup>&</sup>lt;sup>1</sup> Semiconductor properties are well documented in many textbooks; an excellent online resource is provided by the Ioffe Institute of the Russian Academy of Sciences at the web site [1].

6

#### Semiconductors, alloys, heterostructures

**Table 1.2** Properties of some semiconductor lattices: the crystal is D (diamond), ZB (zinc-blende) or W (wurtzite); the gap is D (direct) or I (indirect);  $\epsilon_{\parallel}$  is along the *c* axis,  $\epsilon_{\perp}$  is orthogonal to the *c* axis for wurtzite materials. Permittivities are static to RF. Properties are at 300 K.

Material	Crystal	$E_g$ (eV)	D/I gap	$\epsilon_r$ or $\epsilon_{\parallel}$	$\epsilon_{\perp}$	a (Å)	с (Å)	Density, ρ (g/cm <sup>3</sup> )
С	D	5.50	Ι	5.57		3.57		3.51
Si	D	1.12	Ι	11.9		5.43		2.33
SiC	ZB	2.42	Ι	9.72		4.36		3.17
Ge	D	0.66	Ι	16.2		5.66		5.32
GaAs	ZB	1.42	D	13.2		5.68		5.32
GaP	ZB	2.27	Ι	11.11		5.45		4.14
GaSb	ZB	0.75	D	15.7		6.09		5.61
InP	ZB	1.34	D	12.56		5.87		4.81
InAs	ZB	0.36	D	15.15		6.06		5.67
InSb	ZB	0.23	D	16.8		6.48		5.77
AlP	ZB	2.45	Ι	9.8		5.46		2.40
AlAs	ZB	2.17	Ι	10.06		5.66		3.76
AlSb	ZB	1.62	Ι	12.04		6.13		4.26
CdTe	ZB	1.47	D	10.2		6.48		5.87
GaN	W	3.44	D	10.4	9.5	3.17	5.16	6.09
AlN	W	6.20	D	9.14		3.11	4.98	3.25
InN	W	1.89	D	14.4	13.1	3.54	5.70	6.81
ZnO	W	3.44	D	8.75	7.8	3.25	5.21	5.67



Figure 1.7 The hexagonal wurtzite cell. The *c*-axis corresponds to the direction of the  $\underline{a}_3 = \underline{c}$  vector.

# 1.2.3 Ferroelectric crystals

Ferroelectric materials have a residual spontaneous dielectric polarization after the applied electric field has been switched off. The behavior of such materials is somewhat similar to that of ferromagnetic materials. Below a transition temperature, called the Curie temperature  $T_c$ , ferroelectric materials possess a spontaneous polarization or electric dipole moment. The magnitude of the spontaneous polarization is greatest at temperatures well below the Curie temperature, and approaches zero as the Curie

#### 1.2 Semiconductor crystal structure

7

**Table 1.3** Properties of some ferroelectric crystals. KDP stands for potassium dihydrogen phosphate.

 Data from [2], Ch. 13, Table 2.

Material class	Material	Curie temperature $T_c$ (K)	Spontaneous polarization $P_s$ ( $\mu$ C/cm <sup>2</sup> )
KDP	KH <sub>2</sub> PO <sub>4</sub>	123	4.75
Perovskites	BaTiO <sub>3</sub>	408	26
Perovskites	LiNbO <sub>3</sub>	1480	71
Perovskites	KNbO <sub>3</sub>	708	30

temperature is approached. Ferroelectric materials are inherently piezoelectric; that is, in response to an applied mechanical load, the material will produce an electric charge proportional to the load. Similarly, the material will produce a mechanical deformation in response to an applied voltage. In optoelectronics, ferroelectric materials are particularly important because of the excellent *electrooptic properties*, i.e., the strong variation of the material refractive index with an applied electric field. The crystal structure is often cubic face-centered, and the material is anisotropic and uniaxial. The most important ferroelectric crystal for optical applications is probably lithium niobate, LiNbO3 (LN for short); some other materials (such as barium titanate) belonging to the socalled *perovskite* class are also sometimes used. The crystal structure of perovskites is face-centered cubic. Above the Curie temperature, the crystal is strictly cubic, and positive and negative ions are located in the cell so as to lead to zero dipole moment. Below the Curie temperature, however, a transition takes place whereby positive and negative ions undergo a shift in opposite directions; the crystal structure becomes tetragonal (i.e., the elementary cell height  $a_3$  is different from the basis  $a_1 = a_2$ ) and, due to the charge displacement, a net dipole moment arises. Table 1.3 shows a few properties of ferroelectric crystals, namely the spontaneous polarization  $P_s$  and the Curie temperature [2].

# 1.2.4 Crystal defects

In practice, the crystal lattice is affected by defects, either native (i.e., not involving external atoms) or related to nonnative impurities. Moreover, defects can be point defects (0D), line defects (1D), surface defects (2D), such as dislocations, and volume defects (3D), such as precipitates. Native point defects are *vacancies*, see Fig. 1.8, and *self-interstitials*, while *interstitials* are nonnative atoms placed in the empty space between the already existing lattice atoms. *Substitutional* defects involve an external atom, e.g., a dopant, which replaces one native atom. Typically, dopants act as donors or acceptors only if they are in a substitutional site; if they are in an interstitial site, they are inactive (chemically inactivated).<sup>2</sup>

 $<sup>^2</sup>$  Dopants can also be electrically inactivated when they are not ionized.

Cambridge University Press & Assessment 978-0-521-76344-8 — Semiconductor Devices for High-Speed Optoelectronics Giovanni Ghione Excerpt More Information

## 8 Semiconductors, alloys, heterostructures



Figure 1.8 Point defects in a crystal (1D) and dislocations (2D).

# 1.3 Semiconductor electronic properties

## 1.3.1 The energy–momentum dispersion relation

A crystal is a periodic arrangement of atoms; since each positively charged nucleus induces a spherically symmetric Coulomb potential, superposition yields in total a periodic potential  $U(\underline{r})$  such as

$$U(\underline{r}) = U(\underline{r} + \underline{L}),$$

where  $\underline{L} = k\underline{a}_1 + l\underline{a}_2 + m\underline{a}_3$ . In such a periodic potential, electrons follow the rules of quantum mechanics, i.e., they are described by a set of *wavefunctions* associated with allowed electron states. Allowed states correspond to allowed energy bands, which collapse into energy levels for isolated atoms; allowed bands are separated by forbidden bands. Low-energy electrons are bound to atoms, and only the two topmost allowed bands (the last, being almost full, is the *valence band*; the uppermost, almost empty, is the *conduction band*) take part in carrier transport. As already recalled, the vacuum level  $U_0$  is the minimum energy of an electron free to move in and out of the crystal.

Electrons in a crystal are characterized by an energy-momentum relation  $E(\underline{k})$ , where the wavevector  $\underline{k}$  is related to the electron momentum  $\underline{p}$  as  $\underline{p} = \hbar \underline{k}$ . The *dispersion relation*  $E(\underline{k})$  is defined in the  $\underline{k}$  space, also called the reciprocal space; it is generally a multivalued function, periodic in the reciprocal space, whose fundamental period is called the *first Brillouin zone* (FBZ). A number of branches of the dispersion relation refer to the valence band, a number to the conduction band; the total number of branches depends on the crystal structure and is quite large (e.g., 12 for the conduction band and 8 for the valence band) in wurtzite semiconductors.

In cubic semiconductors, the FBZ is a solid with six square faces and eight hexagonal faces, as shown in Fig. 1.9. Owing to symmetries, only a portion of the FBZ, called the *irreducible wedge*, actually includes independent information; all the rest can be recovered by symmetry. Important points in the FBZ are the center ( $\Gamma$  point), the X point (center of the square face), and the L point (center of the hexagonal face).

The full details of the dispersion relation are not essential for understanding lowenergy phenomena in semiconductors; attention can be restricted to the branches Cambridge University Press & Assessment 978-0-521-76344-8 — Semiconductor Devices for High-Speed Optoelectronics Giovanni Ghione Excerpt More Information

#### 1.3 Semiconductor electronic properties

9



Figure 1.9 The first Brillouin zone (FBZ) in a cubic lattice (lattice constant *a*).



Figure 1.10 Simplified dispersion relation for GaAs.

describing low-energy electrons in the conduction band (around the *conduction band edge*  $E_c$ ) and high-energy electrons (low-energy holes) in the valence band (around the *valence band edge*  $E_v$ ). Valence band electrons are more efficiently described in terms of pseudoparticles (the *holes*) related to electrons missing from the valence band. Holes behave as particles with positive charge and potential energy opposite to the electron energy, so that the topmost branches of the dispersion relation (i.e., the branches describing low-energy holes) define the valence band edge.

As a relevant example, let us discuss the dispersion relation for a direct-bandgap semiconductor, GaAs. The term *direct bandgap* refers to the fact that the minimum of the conduction band and the maximum of the valence band (both located in the  $\Gamma$  point) correspond to the same momentum  $\hbar \underline{k}$ , in this case  $\hbar \underline{k} = 0$ . The dispersion relation shown in Fig. 1.10 is simplified, in the sense that only the lowest branch of the conduction band is shown, while three branches of the valence band appear, the *heavy hole* (HH), the *light hole* (LH), and the *split-off* band. Light and heavy hole bands are degenerate, i.e., they share the same minimum in the  $\Gamma$  point, and they differ because of the *E*( $\underline{k}$ ) curvature near the minimum, which corresponds to a larger or smaller hole effective mass. The split-off band enters some transport and optical processes but can be neglected in a first-order treatment. The conduction band has the lowest minimum at

10

#### Semiconductors, alloys, heterostructures

the  $\Gamma$  point, and two secondary minima at the *L* and *X* points. The main gap is 1.42 eV, while the secondary gaps are 1.72 eV (*L* point) and 1.90 eV (*X* point). Only a section of the dispersion relation is presented, running from the *L* point to the  $\Gamma$  point (the center of the FBZ), and then from the  $\Gamma$  point to the *X* point and back to the origin through the *K* point.

Since electrons and holes have, at least in the absence of an applied field, a Boltzmann energy distribution (i.e., their probability to have energy *E* is proportional to  $\exp(-E/k_BT)$ , where  $k_BT = 26$  meV at ambient temperature), most electrons and holes can be found close to the conduction band and valence band edges, respectively.

Consider now the lowest minimum of the conduction band or highest maximum in the valence band; the dispersion relation can be approximated (around the  $\Gamma$  point) by a parabola as

$$E_n - E_c \approx \frac{\hbar^2 k^2}{2m_n^*}, \quad E_v - E_h \approx \frac{\hbar^2 k^2}{2m_h^*},$$

where  $m_n^*$  and  $m_h^*$  are the electron and hole *effective masses.*<sup>3</sup> Therefore, the electron kinetic energy  $E_n - E_c$  or hole kinetic energy  $E_v - E_h$  (assuming the valence band edge energy  $E_v$  and the conduction band edge energy  $E_c$  to be the energy of a hole or of an electron, respectively, at rest) have, approximately, the same expression as the free-space particle kinetic energy, but with an effective mass  $m_n^*$  or  $m_h^*$  instead of the *in vacuo* inertial mass  $m_0$ . If the minimum is not located in the center of the first BZ (as for the conduction band of indirect bandgap semiconductors) the momentum (in a dynamic sense) can be defined "with respect to the minimum," so that the following approximation applies:

$$E_n - E_c \approx \frac{\hbar^2 \left| \underline{k} - \underline{k}_{\min} \right|^2}{2m_n^*}.$$

The effective mass can be evaluated from the inverse of the curvature of the dispersion relation around a minimum or a maximum. In general, the approximating surface can be expressed as

$$E_n - E_c = \frac{\hbar^2 k_a^2}{2m_{na}^*} + \frac{\hbar^2 k_b^2}{2m_{nb}^*} + \frac{\hbar^2 k_c^2}{2m_{nc}^*}$$

which is an ellipsoid; the coordinate system coincides with the principal axes. If the three effective masses are equal, the ellipsoid degenerates into a spherical surface, and we say that the minimum is *spherical*, with *isotropic* effective mass. This typically happens at  $\Gamma$  point minima. In indirect-bandgap semiconductors, the constant-energy

$$E_k \left( 1 + \alpha E_k \right) = \frac{\hbar^2 k^2}{2m_n^*},$$

where  $E_k$  is the electron kinetic energy  $E_n - E_c$  and  $\alpha$  is a nonparabolicity correction factor.

<sup>&</sup>lt;sup>3</sup> Corrections to the parabolic approximation accounting for nonparaboliticity effects can be introduced (e.g., in the conduction band) through the expression: