Chapter 17
Periodic Structures

17.1 Introduction

The chapter outlines a number of concepts that are useful in the description of periodic structures. The first sections describe the geometrical entities (characteristic vectors, direct and reciprocal lattices, translation vectors, cells, and Brillouin zones) used for describing a lattice. The analysis focuses on the lattices of cubic type, because silicon and other semiconductor materials used in the fabrication of integrated circuits have this type of structure. The next sections introduce the mathematical apparatus necessary for solving the Schrödinger equation within a periodic structure, specifically, the translation operators, Bloch theorem, and periodic boundary conditions. Basing on such tools, the Schrödinger equation is solved, leading to the dispersion relation of the electrons, whose properties are worked out and discussed. The analogy between a wave packet in a periodic potential and in free space is also outlined. Then, the parabolic-band approximation is introduced, leading to the concept of effective mass and to the explicit calculation of the density of states. Examples of the structure of the conduction and valence bands of silicon, germanium, and gallium arsenide are provided. Considering the importance of two-dimensional and one-dimensional structures in modern technological applications, a detailed derivation of the subbands and the corresponding density of states is also given. Then, the same mathematical concepts used for solving the Schrödinger equation (Bloch theorem and periodic boundary conditions) are applied to the calculation of the lattice's vibrational spectrum in the harmonic approximation. The properties of the eigenvalues and eigenfunctions of the problem are worked out, leading to the expression of the vibrational modes. The complements provide a number of details about the description of crystal planes and directions, and about the connection between the symmetries of the Hamiltonian operator and the properties of its eigenvalues. A number of examples of application of the concepts outlined in the chapter are given in the last part of the complements; specifically, the Kronig–Penney model, showing a one-dimensional calculation of the electrons' dispersion relation, and the derivation of the dispersion relation of the one-dimensional monatomic and diatomic chains. The complements are concluded by a discussion about some analogies between the energy of the electromagnetic field and that of the vibrating lattice, and between the dispersion relation of the electrons and that of the lattice.

M. Rudan, *Physics of Semiconductor Devices,*
DOI 10.1007/978-1-4939-1151-6_17

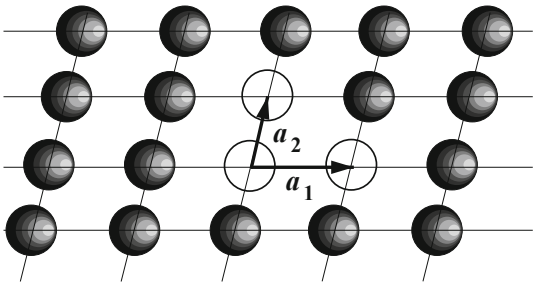

Fig. 17.1 Schematic description of a two-dimensional Bravais lattice of the oblique type. Three atoms have been removed to better show the characteristic vectors. The latter are not orthogonal to each other, and their lengths are different

17.2 Bravais Lattice

The concepts illustrated in the previous chapters will now be applied to study the properties of a specific class of materials, the *crystals*. In fact, in the majority of cases the solid-state devices are manufactured using a crystal as basic material.[1] A crystal is made of a periodic arrangement of an atom, or a group of atoms, called *basis*. As a periodic structure is unlimited in all spatial directions, the general properties of crystals are derived using the provisional hypothesis that the material extends to infinity. The more realistic case of a finite crystal is considered at a later stage.

To describe the properties of a crystal it is convenient to superimpose to it a geometric structure, called *Bravais lattice*, made of an infinite array of discrete points generated by translation operations of the form [10]

$$\mathbf{l} = m_1\,\mathbf{a}_1 + m_2\,\mathbf{a}_2 + m_3\,\mathbf{a}_3, \tag{17.1}$$

with m_1, m_2, m_3 any integers. In (17.1), $\mathbf{l}$ is called *translation vector* while $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ are the *characteristic vectors*; the discrete points generated by (17.1) are called *nodes*. The set of vectors $\mathbf{l}$ is closed under vector addition. Although the characteristic vectors are not necessarily of equal length, nor are they orthogonal to each other, they form a complete set; it follows that any vector $\mathbf{r}$ of the three-dimensional space is expressible as

$$\mathbf{r} = \mu_1\,\mathbf{a}_1 + \mu_2\,\mathbf{a}_2 + \mu_3\,\mathbf{a}_3, \tag{17.2}$$

with μ_1, μ_2, μ_3 real numbers. In a zero-dimensional or one-dimensional space only one Bravais lattice is possible. In a two-dimensional space there are five Bravais lattices, respectively called *oblique*, *rectangular*, *centered rectangular (rhombic)*, *hexagonal*, and *square* [63]. An example of oblique lattice is shown in Fig. 17.1.

[1] Some important exceptions exist. *Thin-Film Transistors* (TFT), commonly used in flat-screen or liquid-crystal displays, are obtained by depositing a semiconductor layer (typically, silicon) over a non-conducting substrate; due to the deposition process, the structure of the semiconductor layer is amorphous or polycrystalline. *Phase-Change Memories* (PCM) exploit the property of specific materials like chalcogenides (for example, $Ge_2Sb_2Te_5$), that switch from the crystalline to the amorphous state, and vice versa, in a controlled way when subjected to a suitable electric pulse.

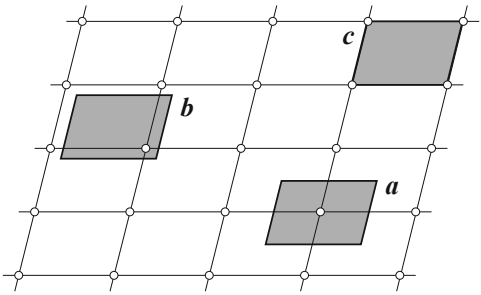

Fig. 17.2 Examples of cells in a two-dimensional Bravais lattice of the oblique type

Another important concept is that of *cell*. Still considering a two-dimensional lattice of the oblique type, a cell is a two-dimensional surface that is associated to each node and has the following properties: (*i*) the association is one-to-one, (*ii*) the cells are equal to each other, and (*iii*) the union of the cells covers the lattice exactly. The cell so defined is not unique; by way of examples, all the shaded surfaces shown in Fig. 17.2 fulfill the properties listed above. It may seem that the cell of case (*c*) is not correct, because it touches four nodes; however, each node is shared by four cells, so that the one-to-one correspondence is maintained. In fact, the type of cell shown in case (*c*) is most useful to extend the definitions given so far to the more realistic three-dimensional case. One notes in passing that the common value of the area of the cells depicted in Fig. 17.2 is $A = |\mathbf{a}_1 \wedge \mathbf{a}_2|$, with $\mathbf{a}_1$, $\mathbf{a}_2$ the characteristic vectors indicated in Fig. 17.1.

In a three-dimensional space the simplest cells are three-dimensional volumes, that fulfill the same properties as in the two-dimensional case and have the atoms placed at the corners. Such an arrangement is called *primitive*. It can be shown that seven primitive arrangements are possible; each of them may further by enriched in five possible ways, by adding (*a*) one atom at the center of the cell, or (*b*) one atom at the center of each face of the cell, or (*c*) one atom at the center of each pair of cell faces (this can be done in three different ways). The addition of extra atoms to the primitive arrangement is called *centering*. The total number of three-dimensional arrangements, including the primitive ones, is thus $7 \times 6 = 42$; however, it is found that not all of them are distinct, so that the actual number of distinct three-dimensional cells reduces to 14 [63].

A three-dimensional lattice whose characteristic vectors are mutually orthogonal and of equal length is called *cubic*. Besides the primitive one, other two arrangements of cubic type are possible: the first one has one additional atom at the cell's center and is called *body-centered cubic* (BCC), the second one has one additional atom at the center of each cell's face and is called *face-centered cubic* (FCC). A portion of a lattice of the FCC type is shown in Fig. 17.3. Examples of chemical species whose crystalline phase has a cubic cell are carbon (C) in the diamond phase, silicon (Si), and germanium (Ge); in fact, this type of crystallization state is also collectively indicated with *diamond structure*. Elements like C, Si, and Ge belong to the fourth column of the periodic table of elements; Si and Ge are semiconductors, while C in

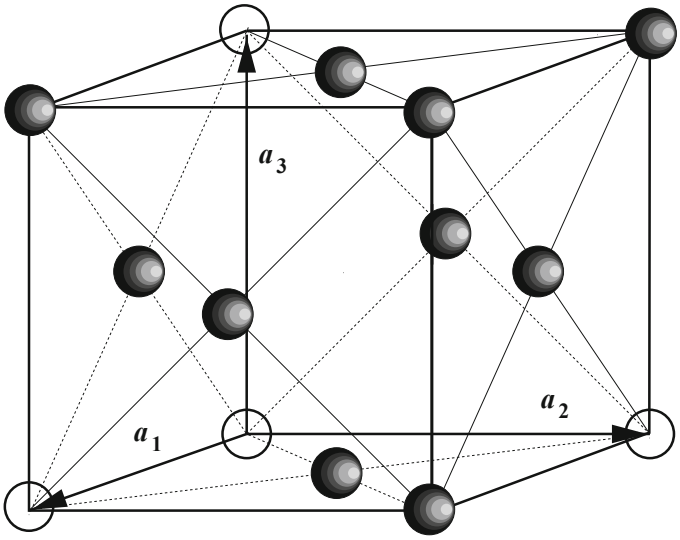

Fig. 17.3 Schematic description of a three-dimensional Bravais lattice of the FCC type. Four atoms have been removed to better show the characteristic vectors. The latter are orthogonal to each other, and of equal length

Table 17.1 Crystal constants of silicon and germanium

Material	Lattice constant (nm)	Interatomic distance (nm)
Si	0.543	0.233
Ge	0.566	0.244

the diamond phase is an insulator.[2] The FCC cell is also exhibited by some compound materials, an example of which is gallium arsenide (GaAs), a semiconductor of the so-called *III-V type*.[3] Some properties of GaAs, along with those of other III-V semiconductors, are listed in Tables 17.1 and 17.2. The type of crystallization state of GaAs and the other materials of the table is called *zincblende structure*.[4]

[2] The meaning of "insulator" or "semiconductor", as well as that of "conductor", is specified in Sect. 18.2.

[3] The term "III-V" derives from the fact the Ga and As belong, respectively, to the third and fifth column of the periodic table of elements.

[4] *Zincblende* is another name for *Sphalerite*, an economically important mineral whence zinc is extracted. It consists of zinc sulphide in crystalline form with some contents of iron, (Zn,Fe)S.

Table 17.2 Crystal constants of some III-V semiconductors

Material	Lattice constant (nm)	Interatomic distance (nm)
GaAs	0.563	0.244
GaP	0.545	0.236
GaSb	0.609	0.264
InAs	0.605	0.262
InP	0.586	0.254
InSb	0.647	0.280
AlSb	0.613	0.265

17.3 Reciprocal Lattice

As indicated in Sect. 17.2, the type of cell having the nodes at the corners is the most useful one for introducing a number of definitions. For instance, observing that the characteristic vectors coincide with the cell's edges (this property is evident for the FCC cell as shown in Fig. 17.3, however, it applies also to the other types of cells), one obtains for the cell's volume

$$\tau_l = \mathbf{a}_1 \cdot \mathbf{a}_2 \wedge \mathbf{a}_3 = \mathbf{a}_2 \cdot \mathbf{a}_3 \wedge \mathbf{a}_1 = \mathbf{a}_3 \cdot \mathbf{a}_1 \wedge \mathbf{a}_2, \tag{17.3}$$

where the orientation of the characteristic vectors is chosen in such a way as to make τ_l positive.

The major advantage of dealing with a periodic structure is the possibility of using the Fourier series. In fact, the functions describing the physical properties of the structure are expected to be periodic, so that one can take advantage of the Fourier expansion. The latter, in turn, entails a transformation from the coordinate space to another space reciprocal to it. Considering, for instance, the one-dimensional case of the Fourier transform given in (C.17), there the k space is the reciprocal of the x space; similarly, in an n-dimensional space like that considered in (C.20), the $\mathbf{k}$ space is the reciprocal of the $\mathbf{x}$ space: both vectors $\mathbf{k}$, $\mathbf{x}$ are linear combinations of the same set of mutually-orthogonal, unit vectors $\mathbf{i}_1, \ldots, \mathbf{i}_n$, so that the scalar product $\mathbf{k} \cdot \mathbf{x}$ yields $k_1 x_1 + \cdots + k_n x_n$. In the case considered here, instead, one must account for the fact that the reference $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ of the space under consideration is made of vectors that, in general, are neither mutually orthogonal nor of equal length. For this reason it is necessary to introduce a *reciprocal lattice* by defining its characteristic vectors as

$$\mathbf{b}_1 = \frac{\mathbf{a}_2 \wedge \mathbf{a}_3}{\tau_l}, \qquad \mathbf{b}_2 = \frac{\mathbf{a}_3 \wedge \mathbf{a}_1}{\tau_l}, \qquad \mathbf{b}_3 = \frac{\mathbf{a}_1 \wedge \mathbf{a}_2}{\tau_l}. \tag{17.4}$$

From the definition (17.3) of τ_l and the mixed-product property (A.32) one finds that the characteristic vectors fulfill the orthogonality and normalization relation

$$\mathbf{a}_i \cdot \mathbf{b}_j = \delta_{ij}, \tag{17.5}$$

with δ_{ij} the Kronecker symbol (A.18). The translation vectors and the general vectors of the reciprocal lattice are linear combinations of $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ whose coefficients are, respectively, integer numbers or real numbers. To distinguish the lattice based on $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ from the one based on $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, the latter is also called *direct lattice*. The common value of the cells' volume in the reciprocal lattice is

$$\tau_G = \mathbf{b}_1 \cdot \mathbf{b}_2 \wedge \mathbf{b}_3 = \mathbf{b}_2 \cdot \mathbf{b}_3 \wedge \mathbf{b}_1 = \mathbf{b}_3 \cdot \mathbf{b}_1 \wedge \mathbf{b}_2. \tag{17.6}$$

Observing that $\mathbf{a}_i$ is a length, from (17.4) it follows that $\mathbf{b}_j$ is the inverse of a length; as a consequence, the units of τ_G are m^{-3}, and the product $\gamma = \tau_l\, \tau_G$ is a dimensionless constant. The value of γ is found by combining the first relation in (17.4), which yields $\tau_l\, \mathbf{b}_1 = \mathbf{a}_2 \wedge \mathbf{a}_3$, with (17.6), so that

$$\gamma = \tau_l\, \tau_G = \tau_l\, \mathbf{b}_1 \cdot \mathbf{b}_2 \wedge \mathbf{b}_3 = \mathbf{a}_2 \wedge \mathbf{a}_3 \cdot (\mathbf{b}_2 \wedge \mathbf{b}_3) = \mathbf{a}_2 \cdot \mathbf{a}_3 \wedge (\mathbf{b}_2 \wedge \mathbf{b}_3)\,, \tag{17.7}$$

where the last equality is due to the invariance of the mixed product upon interchange of the "wedge" and "dot" symbols (Sect. A.7). Then, using (A.33) to resolve the double vector product,

$$\tau_l\, \tau_G = \mathbf{a}_2 \cdot (\mathbf{a}_3 \cdot \mathbf{b}_3\, \mathbf{b}_2 - \mathbf{a}_3 \cdot \mathbf{b}_2\, \mathbf{b}_3) = \mathbf{a}_2 \cdot (\delta_{33}\, \mathbf{b}_2 - \delta_{32}\, \mathbf{b}_3) = \mathbf{a}_2 \cdot \mathbf{b}_2 = 1. \tag{17.8}$$

Also, after defining $\mathbf{e}_1 = \mathbf{b}_2 \wedge \mathbf{b}_3 / \tau_G$ one finds, by a similar calculation,

$$\mathbf{e}_1 = \frac{\mathbf{b}_2 \wedge \mathbf{b}_3}{\tau_G} = \frac{\mathbf{b}_2 \wedge (\mathbf{a}_1 \wedge \mathbf{a}_2)}{\tau_l\, \tau_G} = \mathbf{b}_2 \cdot \mathbf{a}_2\, \mathbf{a}_1 - \mathbf{b}_2 \cdot \mathbf{a}_1\, \mathbf{a}_2 = \mathbf{a}_1. \tag{17.9}$$

From the properties that provide the definition of cell (Sect. 17.2) it follows that, for a given lattice, the volume of the cell does not depend on its form. It follows that $\tau_l\, \tau_G = 1$ no matter what the cell's form is and, from (17.9), that the direct lattice is the reciprocal of the reciprocal lattice.

Given a direct lattice of characteristic vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, it is convenient to introduce, besides the reciprocal lattice of characteristic vectors $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ defined by (17.4), another lattice called *scaled reciprocal lattice*, whose characteristic vectors are $2\,\pi\, \mathbf{b}_1$, $2\,\pi\, \mathbf{b}_2$, $2\,\pi\, \mathbf{b}_3$. A translation vector of the scaled reciprocal lattice has the form

$$\mathbf{g} = n_1\, 2\,\pi\, \mathbf{b}_1 + n_2\, 2\,\pi\, \mathbf{b}_2 + n_3\, 2\,\pi\, \mathbf{b}_3, \tag{17.10}$$

with n_1, n_2, n_3 any integers, whereas a general vector of the scaled reciprocal lattice has the form

$$\mathbf{k} = \nu_1\, 2\,\pi\, \mathbf{b}_1 + \nu_2\, 2\,\pi\, \mathbf{b}_2 + \nu_3\, 2\,\pi\, \mathbf{b}_3, \tag{17.11}$$

with ν_1, ν_2, ν_3 any real numbers. From the definitions (17.1, 17.10) of $\mathbf{l}$ and $\mathbf{g}$ one finds

$$\mathbf{l} \cdot \mathbf{g} = \sum_{is=1}^{3} m_i\, \mathbf{a}_i \cdot n_s\, 2\,\pi\, \mathbf{b}_s = 2\,\pi \sum_{is=1}^{3} m_i\, n_s\, \delta_{is} = 2\,\pi \sum_{i=1}^{3} m_i\, n_i, \tag{17.12}$$

namely, $\mathbf{l} \cdot \mathbf{g} = 2\pi M$ with M an integer. It follows that

$$\exp\left[\mathrm{i}\,\mathbf{g} \cdot (\mathbf{r} + \mathbf{l})\right] = \exp(\mathrm{i}\,\mathbf{g} \cdot \mathbf{r}) \exp(\mathrm{i}\,2\pi M) = \exp(\mathrm{i}\,\mathbf{g} \cdot \mathbf{r}), \tag{17.13}$$

that is, $\exp(\mathrm{i}\,\mathbf{g} \cdot \mathbf{r})$ is periodic in the $\mathbf{r}$ space. This shows the usefulness of the scaled reciprocal lattice for treating problems related to periodic structures. In fact, given a periodic function in the $\mathbf{r}$ space, $F(\mathbf{r}+\mathbf{l}) = F(\mathbf{r})$, the Fourier expansion is generalized to the non-orthogonal case as

$$F(\mathbf{r}) = \sum_{\mathbf{g}} F_{\mathbf{g}} \exp(\mathrm{i}\,\mathbf{g} \cdot \mathbf{r}), \qquad F_{\mathbf{g}} = \frac{1}{\tau_l} \int_{\tau_l} F(\mathbf{r}) \exp(-\mathrm{i}\,\mathbf{g} \cdot \mathbf{r})\, \mathrm{d}^3 r, \tag{17.14}$$

with $\sum_{\mathbf{g}} = \sum_{n_1} \sum_{n_2} \sum_{n_3}$. The property holds also in reverse, namely,

$$\exp[\mathrm{i}\,\mathbf{l} \cdot (\mathbf{k} + \mathbf{g})] = \exp(\mathrm{i}\,\mathbf{l} \cdot \mathbf{k}), \tag{17.15}$$

so that, given a periodic function in the $\mathbf{k}$ space, $\Phi(\mathbf{k} + \mathbf{g}) = \Phi(\mathbf{k})$, the following expansion holds:

$$\Phi(\mathbf{k}) = \sum_{\mathbf{l}} \Phi_{\mathbf{l}} \exp(\mathrm{i}\,\mathbf{l} \cdot \mathbf{k}), \qquad \Phi_{\mathbf{l}} = \frac{1}{\tau_g} \int_{\tau_g} \Phi(\mathbf{k}) \exp(-\mathrm{i}\,\mathbf{l} \cdot \mathbf{k})\, \mathrm{d}^3 k, \tag{17.16}$$

with $\sum_{\mathbf{l}} = \sum_{m_1} \sum_{m_2} \sum_{m_3}$. From (17.8) one also finds that in the scaled reciprocal lattice the volume of the cell is given by

$$\tau_g = (2\pi)^3 \tau_G = \frac{(2\pi)^3}{\tau_l}. \tag{17.17}$$

The origin of the reference of the direct or reciprocal space has not been identified so far. After selecting the origin, consider the cell of the $\mathbf{r}$ space whose sides emanate from it; these sides are made to coincide with the characteristic vectors (compare, e.g., with Fig. 17.3), so that, to any point $\mathbf{r} = \mu_1 \mathbf{a}_1 + \mu_2 \mathbf{a}_2 + \mu_3 \mathbf{a}_3$ that belongs to the interior or the boundary of the cell, the following restriction apply: $0 \le \mu_i \le 1$. Similarly, if one considers the cell of the $\mathbf{k}$ space whose sides emanate from the origin, for any point $\mathbf{k} = \nu_1 2\pi \mathbf{b}_1 + \nu_2 2\pi \mathbf{b}_2 + \nu_3 2\pi \mathbf{b}_3$ that belongs to the interior or the boundary of the cell it is $0 \le \nu_i \le 1$.

17.4 Wigner–Seitz Cell—Brillouin Zone

It has been mentioned in Sect. 17.2 that the cell properties do not identify the cell uniquely. Cells of a form different from that shown, e.g., in the two-dimensional example of Fig. 17.2 can be constructed. Among them, of particular interest is the *Wigner–Seitz* cell, whose construction is shown in Fig. 17.4 still considering a two-dimensional lattice of the oblique type. First, the node to be associated to the cell is connected to its nearest neighbors as shown in the figure (the connecting segments are

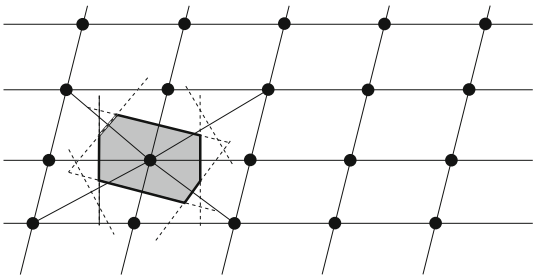

Fig. 17.4 A Wigner–Seitz cell in a two-dimensional, oblique lattice

the continuous lines); then, the axis of each segment is drawn (dashed lines). As the axis is the locus of points having the same distance from the extrema of the connecting segment, one can select a contour, made of portions of the axes (thick line), such that the following property holds: any point inside the contour is closer to the node associated to the cell than to any other lattice node. From the construction described above it also follows that the shaded surface so obtained fulfills the requisites listed in Sect. 17.2, so it is a cell proper; such a cell, named after Wigner and Seitz, fulfills the additional property of the closeness of the internal points. Its construction in the three-dimensional case is similar, the only difference being that the axis is replaced with the plane normal to the connecting segment at the midpoint.

The concept of Wigner–Seitz cell, that has been introduced above using the direct lattice by way of example, is applicable also to the reciprocal and scaled-reciprocal lattices. In the scaled reciprocal lattice, the Wigner-Seitz cell associated to the origin is called *first Brillouin zone*. The set of Wigner-Seitz cells adjacent to the first Brilluoin zone is called *second Brillouin zone*, and so on. The first Brillouin zone of the FCC lattice is shown in Fig. 17.5; its boundary is made of six square faces and eight hexagonal faces. The center of the zone is called *Γ point*; the k_1 axis belongs to the [100] crystal direction[5] and intercepts the boundary of the Brilluoin zone at two opposite positions called *X points*, that coincide with the center of square faces; the k_2, k_3 axes belong to the [010] and [001] directions, respectively, and intercept the boundary at X points as well. In turn, the {111} directions intercept the boundary at positions called *L points*, that coincide with the center of the hexagonal faces. There is a total of eight L points, because the set {111} is made of the [111], [$\bar{1}$11], [$\bar{1}\bar{1}$1], [1$\bar{1}$1] directions along with those of complementary signs.

17.5 Translation Operators

The typical procedure by which the physical properties of periodic structures, like crystals, are investigated, entails the solution of eigenvalue equations generated by quantum-mechanical operators. In this respect it is useful to introduce a class of

[5] The symbols indicating the crystal directions are illustrated in Sect. 17.8.1.

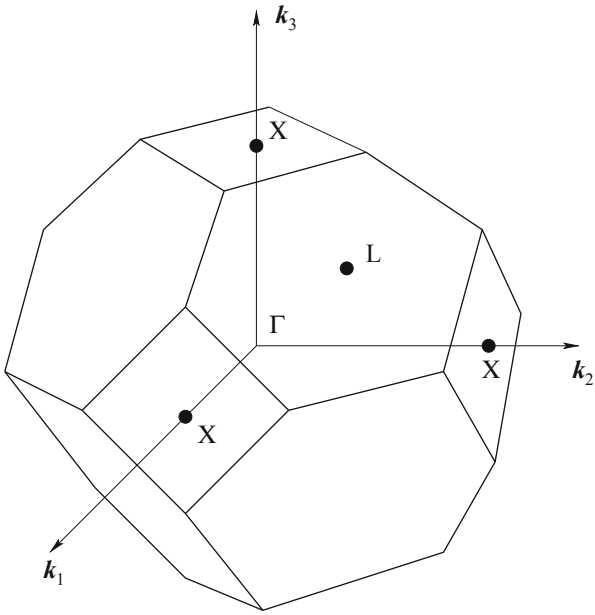

Fig. 17.5 The first Brillouin zone of the FCC lattice

operators, called *translation operators*, that are associated to the direct-lattice vectors $\mathbf{l}$ defined by (17.1). To ease the notation, the translation operator associated to the ith lattice vector $\mathbf{l}_i$ is indicated with $\mathcal{T}_i = \mathcal{T}(\mathbf{l}_i)$. A translation operator is defined by the property

$$\mathcal{T}_i f(\mathbf{r}) = f(\mathbf{r} + \mathbf{l}_i) \tag{17.18}$$

for all functions f defined over the direct lattice. It is easily found that translation operators are linear and non-Hermitean. Also, they commute with each other; in fact,

$$\mathcal{T}_i \mathcal{T}_s f(\mathbf{r}) = f(\mathbf{r} + \mathbf{l}_i + \mathbf{l}_s) = \mathcal{T}_s \mathcal{T}_i f(\mathbf{r}) \tag{17.19}$$

for all functions f and all indices i, s. Remembering the property derived in Sect. 10.4, it follows that all translation operators have a common set of eigenfunctions v. Combining this result with definition (17.18) provides the form of the eigenvalues α of the translation operators. For this, consider the operators associated to three arbitrary vectors $\mathbf{l}_i$, $\mathbf{l}_s$, and $\mathbf{l}_u$, and generate the eigenvalue equations

$$\mathcal{T}_i v = \alpha(\mathbf{l}_i)\, v, \qquad \mathcal{T}_s v = \alpha(\mathbf{l}_s)\, v, \qquad \mathcal{T}_u v = \alpha(\mathbf{l}_u)\, v. \tag{17.20}$$

If one now lets $\mathbf{l}_u = \mathbf{l}_i + \mathbf{l}_s$, it follows $\mathcal{T}_i \mathcal{T}_s v(\mathbf{r}) = v(\mathbf{r} + \mathbf{l}_i + \mathbf{l}_s) = \mathcal{T}_u v(\mathbf{r})$, that is, $\mathcal{T}_i \mathcal{T}_s = \mathcal{T}_u$. On the other hand, the first two eigenvalue equations in (17.20) provide $\mathcal{T}_i \mathcal{T}_s v = \mathcal{T}_i\, \alpha(\mathbf{l}_s)\, v = \alpha(\mathbf{l}_i)\, \alpha(\mathbf{l}_s)\, v$ which, combined with the third one, yields

$$\alpha(\mathbf{l}_i)\, \alpha(\mathbf{l}_s) = \alpha(\mathbf{l}_u) = \alpha(\mathbf{l}_i + \mathbf{l}_s). \tag{17.21}$$

The result shows that the functional dependence of α on the translation vector must be of the exponential type:

$$\alpha(\mathbf{l}) = \exp(\mathbf{c} \cdot \mathbf{l}), \qquad \mathbf{c} = \sum_{s=1}^{3} (\Re\chi_s + \mathrm{i}\,\Im\chi_s)\, 2\pi\, \mathbf{b}_s, \tag{17.22}$$

where $\mathbf{c}$ is a complex vector whose units are the inverse of a length. For this reason $\mathbf{c}$ is given the general form shown in the second relation of (17.22), with $\mathbf{b}_s$ a characteristic vector of the reciprocal lattice, $s = 1, 2, 3$, and χ_s a dimensionless complex number that is provisionally left unspecified.

17.5.1 Bloch Theorem

The expression (17.22) of the eigenvalues of the translation operators makes it possible to specify some property of the eigenfunctions; in fact, combining the definition of translation operator, $\mathcal{T} v(\mathbf{r}) = v(\mathbf{r} + \mathbf{l})$, with the eigenvalue equation $\mathcal{T} v(\mathbf{r}) = \exp(\mathbf{c} \cdot \mathbf{l})\, v(\mathbf{r})$ yields

$$v_{\mathbf{c}}(\mathbf{r} + \mathbf{l}) = \exp(\mathbf{c} \cdot \mathbf{l})\, v_{\mathbf{c}}(\mathbf{r}), \tag{17.23}$$

called *Bloch's theorem (first form)*. The importance of this result can be appreciated by observing that, if v is known within a lattice cell, and $\mathbf{c}$ is given, then the eigenfunction can be reconstructed everywhere else. The index in (17.23) reminds one that the eigenfunction depends on the choice of $\mathbf{c}$. The theorem can be recast differently by defining an auxiliary function $u_{\mathbf{c}}(\mathbf{r}) = v_{\mathbf{c}}(\mathbf{r})\, \exp(-\mathbf{c} \cdot \mathbf{r})$, so that

$$v_{\mathbf{c}}(\mathbf{r} + \mathbf{l}) = \exp(\mathbf{c} \cdot \mathbf{l})\, v_{\mathbf{c}}(\mathbf{r}) = \exp(\mathbf{c} \cdot \mathbf{l})\, u_{\mathbf{c}}(\mathbf{r})\, \exp(\mathbf{c} \cdot \mathbf{r}). \tag{17.24}$$

In turn, from the definition of $u_{\mathbf{c}}$ one draws $v_{\mathbf{c}}(\mathbf{r} + \mathbf{l}) = u_{\mathbf{c}}(\mathbf{r} + \mathbf{l})\, \exp[\mathbf{c} \cdot (\mathbf{r} + \mathbf{l})]$ which, combined with (17.24), yields the *Bloch theorem (second form)*:

$$v_{\mathbf{c}}(\mathbf{r}) = u_{\mathbf{c}}(\mathbf{r})\, \exp(\mathbf{c} \cdot \mathbf{r}), \qquad u_{\mathbf{c}}(\mathbf{r} + \mathbf{l}) = u_{\mathbf{c}}(\mathbf{r}). \tag{17.25}$$

The second form of Bloch's theorem shows that the auxiliary function $u_{\mathbf{c}}$ is periodic in the direct lattice, so that the eigenfunctions of the translation operators are the product of an exponential function times a function having the lattice periodicity. One notes the similarity of this result with that expressed by (13.32); in fact, the Bloch theorem is a form of the Floquet theorem (Sect. 13.4).

The eigenfunctions of the translation operators play an important role in the description of the physical properties of periodic structures. For this reason, vectors $\mathbf{c}$ of the general form are not acceptable because their real part would make $v_{\mathbf{c}}(\mathbf{r}) = u_{\mathbf{c}}(\mathbf{r})\, \exp(\mathbf{c} \cdot \mathbf{r})$ to diverge as $\mathbf{r}$ departs more and more from the origin.[6]

[6] This aspect is further elaborated in Sect. 17.5.3.

It is then necessary to impose the restriction $\mathbf{c} = \mathrm{i}\,\mathbf{k}$, with $\mathbf{k}$ real. This is achieved by letting $\Re\chi_s = 0$ and $\Im\chi_s = \nu_s$ in the second relation of (17.22), so that the eigenvalues of the translation operators become

$$\alpha(\mathbf{l}) = \exp(\mathrm{i}\,\mathbf{k}\cdot\mathbf{l}), \qquad \mathbf{k} = \sum_{s=1}^{3} \nu_s\, 2\,\pi\, \mathbf{b}_s. \tag{17.26}$$

Remembering (17.15), such eigenvalues are periodic in the scaled reciprocal lattice. In turn, the first and second form of the Bloch theorem become, respectively,

$$v_{\mathbf{k}}(\mathbf{r}+\mathbf{l}) = \exp(\mathrm{i}\,\mathbf{k}\cdot\mathbf{l})\, v_{\mathbf{k}}(\mathbf{r}), \tag{17.27}$$

$$v_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r})\, \exp(\mathrm{i}\,\mathbf{k}\cdot\mathbf{r}), \qquad u_{\mathbf{k}}(\mathbf{r}+\mathbf{l}) = u_{\mathbf{k}}(\mathbf{r}). \tag{17.28}$$

Eigenfunctions of the form (17.27, 17.28) are also called *Bloch functions*. They fulfill the eigenvalue equation $\mathcal{T} v_{\mathbf{k}}(\mathbf{r}) = \exp(\mathrm{i}\,\mathbf{k}\cdot\mathbf{l})\, v_{\mathbf{k}}(\mathbf{r})$ so that, observing that $\mathcal{T}$ is real and taking the conjugate of the eigenvalue equation yields

$$\mathcal{T} v^*_{\mathbf{k}}(\mathbf{r}) = \exp(-\mathrm{i}\,\mathbf{k}\cdot\mathbf{l})\, v^*_{\mathbf{k}}(\mathbf{r}). \tag{17.29}$$

If, instead, one replaces $\mathbf{k}$ with $-\mathbf{k}$ in the original equation, the following is found:

$$\mathcal{T} v_{-\mathbf{k}}(\mathbf{r}) = \exp(-\mathrm{i}\,\mathbf{k}\cdot\mathbf{l})\, v_{-\mathbf{k}}(\mathbf{r}). \tag{17.30}$$

Comparing (17.30) with (17.29) shows that $v^*_{\mathbf{k}}$ and $v_{-\mathbf{k}}$ belong to the same eigenvalue. Moreover, comparing the second expression in (17.26) with (17.11) shows that $\mathbf{k}$ is a vector of the scaled reciprocal lattice.

A further reasoning demonstrates that the variability of $\mathbf{k}$ in (17.27, 17.28) can be limited to a single cell of the scaled reciprocal lattice; for instance, to the first Brillouin zone or, alternatively, to the cell of the $\mathbf{k}$ space whose sides emanate from the origin, so that the coefficients of (17.11) fulfill the relation $0 \le \nu_i \le 1$ as shown in Sect. 17.3. The property derives from the periodicity of the eigenvalues in the scaled reciprocal lattice, $\exp[\mathrm{i}\,(\mathbf{k}+\mathbf{g})\cdot\mathbf{l}] = \exp(\mathrm{i}\,\mathbf{k}\cdot\mathbf{l})$, due to which the values of $\exp(\mathrm{i}\,\mathbf{k}\cdot\mathbf{l})$, with $\mathbf{k}$ ranging over a single cell, provide the whole set of the operator's eigenvalues. As $\exp[\mathrm{i}\,(\mathbf{k}+\mathbf{g})\cdot\mathbf{l}]$ is the same eigenvalue as $\exp(\mathrm{i}\,\mathbf{k}\cdot\mathbf{l})$, the Bloch function $v_{\mathbf{k}+\mathbf{g}}(\mathbf{r})$ is the same as $v_{\mathbf{k}}(\mathbf{r})$. Note that the reasoning does not prevent the eigenvalue from being degenerate: if this is the case, one finds

$$v^{(1)}_{\mathbf{k}+\mathbf{g}} = v^{(1)}_{\mathbf{k}}, \qquad v^{(2)}_{\mathbf{k}+\mathbf{g}} = v^{(2)}_{\mathbf{k}}, \qquad \ldots \tag{17.31}$$

17.5.2 Periodic Operators

An operator $\mathcal{A}$ is *periodic* in the direct lattice if $\mathcal{A}(\mathbf{r}+\mathbf{l}) = \mathcal{A}(\mathbf{r})$ for all vectors $\mathbf{r}$ and all translation vectors $\mathbf{l}$ of the direct lattice. Periodic operators commute with translation operators: this is shown by letting $v' = \mathcal{A}v$, so that

$$\mathcal{T}(\mathbf{l})\mathcal{A}(\mathbf{r})v(\mathbf{r}) = \mathcal{T}(\mathbf{l})v'(\mathbf{r}) = v'(\mathbf{r}+\mathbf{l}) =$$

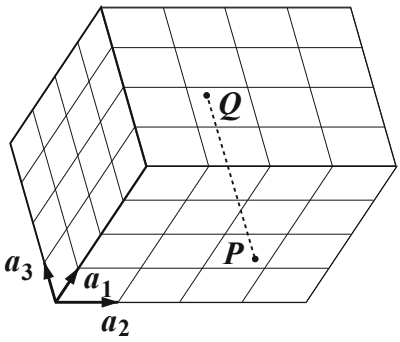

Fig. 17.6 A finite block of material obtained by sectioning a crystal by means of three pairs of parallel crystal planes

$$= \mathcal{A}(\mathbf{r}+\mathbf{l})v(\mathbf{r}+\mathbf{l}) = \mathcal{A}(\mathbf{r})v(\mathbf{r}+\mathbf{l}) = \mathcal{A}(\mathbf{r})\mathcal{T}(\mathbf{l})v(\mathbf{r}). \tag{17.32}$$

From the commutativity property $\mathcal{T}\mathcal{A} = \mathcal{A}\mathcal{T}$ it follows that $\mathcal{T}$ and $\mathcal{A}$ have a common set of eigenfunctions, so that the eigenfunctions of a periodic operator are Bloch functions; letting $A_{\mathbf{k}}$ be the eigenvalue, one has

$$\mathcal{A}v_{\mathbf{k}} = A_{\mathbf{k}}\, v_{\mathbf{k}}, \tag{17.33}$$

with $\mathbf{k}$ belonging to a single cell of the scaled reciprocal lattice, and $v_{\mathbf{k}+\mathbf{g}} = v_{\mathbf{k}}$. Since an eigenfunction belongs to one eigenvalue only, it follows

$$A_{\mathbf{k}+\mathbf{g}} = A_{\mathbf{k}}, \tag{17.34}$$

namely, if an operator is periodic in the direct lattice, its eigenvalues are periodic in the scaled reciprocal lattice.

17.5.3 Periodic Boundary Conditions

In the derivation of the Bloch theorem, carried out in Sect. 17.5.1, it has been observed that complex vectors $\mathbf{c}$ in the expression $v_{\mathbf{c}}(\mathbf{r}) = u_{\mathbf{c}}(\mathbf{r})\ \exp(\mathbf{c}\cdot\mathbf{r})$ of the eigenfunctions are not acceptable because their real part would make the function to diverge. In fact, as noted in Sect. 17.5.2, such eigenfunctions belong also to the operators that commute with the translation operators and may describe physical properties of the crystal, so that diverging solutions must be discarded. On the other hand, such a divergence is due to the assumption that the crystal is unlimited; in the more realistic case of a finite block of material, diverging solutions would not appear. Unfortunately, a finite block of material is not periodic, hence the useful concepts and properties worked out so far in this chapter are not applicable to it.

To further investigate on the subject one notes that a finite block may be thought of as derived from the original crystal by sectioning the latter using three pairs of parallel crystal planes, as shown in Fig. 17.6. One of the vertices of the block coincides with

the origin of the reference of the direct lattice, and the block's sides are aligned with the characteristic vectors. Also, the type of cell chosen here is the one whose sides coincide with the characteristic vectors themselves. The relation between the total number N_c of the block's cells, the block's volume Ω, and that of the cell is easily found to be

$$N_c = N_1\, N_2\, N_3 = \frac{\Omega}{\tau_l}, \tag{17.35}$$

with N_s the number of cell sides that match the side of the block in the sth direction. In a finite volume of material, the number of cells that belong to the interior is typically much larger than the number of cells that are adjacent to the boundaries; when solving a single-electron Schrödinger equation within such a structure, it is found that in the interior of it the eigenfunctions practically coincide with Bloch functions, whereas the effect of the real part of vector $\mathbf{c}$ in $v_{\mathbf{c}}(\mathbf{r}) = u_{\mathbf{c}}(\mathbf{r})\, \exp(\mathbf{c} \cdot \mathbf{r})$ becomes relevant only when the position under investigation is close to a boundary. In fact, the real part of $\mathbf{c}$ is such that the eigenfunctions become vanishingly small far away from the volume considered [15, Sects. F-XI, O-III].

The considerations above show that for practical purposes one may keep the analysis based on the original periodicity of the unlimited structure by replacing the vanishing boundary conditions with a different type of conditions, able to formally restore periodicity in a finite structure. This is accomplished by imposing the identity of the Bloch functions corresponding to two boundary points facing each other along the direction of a characteristic vector. This prescription, called *periodic boundary condition* or *Born–Von Karman boundary condition*, is illustrated with the aid of Fig. 17.6. Consider for instance point $\mathbf{r} = \mu_1\, \mathbf{a}_1 + \mu_2\, \mathbf{a}_2$ (labeled P in the figure), that belongs to the boundary plane defined by $\mathbf{a}_1$, $\mathbf{a}_2$. Point Q facing P on the opposite boundary plane is such that $Q - P = \mathbf{l} = N_3\, \mathbf{a}_3$ whence, applying the first form (17.23) of Bloch's theorem, one obtains $v_c(\mathbf{r} + N_3\, \mathbf{a}_3) = \exp(N_3\, \mathbf{c} \cdot \mathbf{a}_3)\, v_c(\mathbf{r})$. Imposing $v_c(\mathbf{r} + N_3\, \mathbf{a}_3) = v_c(\mathbf{r})$ yields $N_3\, \mathbf{c} \cdot \mathbf{a}_3 = \mathrm{i}\, n_3\, 2\,\pi$, with n_3 any integer, so that, using expression (17.22) for $\mathbf{c}$,

$$N_3\, \mathbf{c} \cdot \mathbf{a}_3 = 2\,\pi\, N_3\, (\Re\chi_3 + \mathrm{i}\, \Im\chi_3) = \mathrm{i}\, n_3\, 2\,\pi. \tag{17.36}$$

In conclusion, $\Re\chi_3 = 0$ and $\Im\chi_3 = 2\,\pi\, n_3/N_3$. The same reasoning is repeated along the other directions, to finally yield

$$\mathbf{c} = \mathrm{i}\, \mathbf{k}, \qquad \mathbf{k} = \sum_{s=1}^{3} \frac{n_s}{N_s}\, 2\,\pi\, \mathbf{b}_s. \tag{17.37}$$

In summary, the application of the periodic boundary conditions gives $\mathbf{c}$ the same imaginary form $\mathbf{c} = \mathrm{i}\, \mathbf{k}$ that was found in an unlimited structure, the difference being that in a finite structure the components of $\mathbf{k}$ are discrete instead of being continuous: given the size of the structure, which is prescribed by N_1, N_2, N_3, each $\mathbf{k}$ vector of the scaled reciprocal lattice is associated to a triad of integers n_1, n_2, n_3.

Note that the reasoning carried out at the end of Sect. 17.5.1 about the variability of $\mathbf{k}$ still holds; as a consequence, $\mathbf{k}$ can be restricted to a single cell of the scaled

reciprocal lattice, so that its coefficients $\nu_s = n_s/N_s$ fulfill the relation $0 \leq \nu_s \leq 1$ as shown in Sect. 17.3. In fact, as $\nu_s = 0$ and $\nu_s = 1$ are redundant, the above relation must more appropriately be recast as $0 \leq n_s/N_s < 1$, corresponding to $n_s = 0, 1, \ldots, N_s - 1$ or, alternatively, $0 < n_s/N_s \leq 1$, corresponding to $n_s = 1, 2, \ldots, N_s$. In both cases, n_s can take N_s distinct values, so that the total number of distinct $\mathbf{k}$ vectors in a cell of the scaled reciprocal lattice is $N_1\, N_2\, N_3 = N_c$. From (17.35) one finds that such a number equals the number of the structure's cells in the direct lattice. Also, as the $\mathbf{k}$ vectors are equally spaced in each direction, their density in the reciprocal scaled lattice is uniform; it is given by the ratio N_c/τ_g, with τ_g the cell's volume. Remembering that the latter is invariant (Sect. 17.3), one may think of $\mathbf{k}$ as restricted to the first Brillouin zone. Combining (17.8) with (17.17) and (17.35) yields for the density

$$\frac{N_c}{\tau_g} = \frac{\Omega/\tau_l}{\tau_g} = \frac{\Omega}{(2\,\pi)^3}. \tag{17.38}$$

One can also define a combined density of the $\mathbf{k}$ vectors in the $\mathbf{r}, \mathbf{k}$ space, which is obtained by dividing (17.38) by the volume Ω. This yields the dimensionless combined density

$$\frac{1}{\Omega}\,\frac{N_c}{\tau_g} = \frac{1}{(2\,\pi)^3}. \tag{17.39}$$

17.6 Schrödinger Equation in a Periodic Lattice

The concepts introduced in the previous sections of this chapter are applied here to the solution of the Schrödinger equation in a periodic lattice. It is assumed provisionally that the lattice is unlimited; as a consequence, the components of the $\mathbf{k}$ vector are continuous. The Schrödinger equation to be solved derives from the single-electron operator (16.27) obtained from the separation procedure outlined in Sects. 16.2–16.5. This means the nuclei are kept fixed and the force acting on the electron derives from a potential energy[7] having the periodicity of the direct lattice: $V(\mathbf{r}+\mathbf{l}) = V(\mathbf{r})$, with $\mathbf{l}$ given by (17.1). As mentioned in Sects. 16.4 and 16.5, the external forces are absent ($U_{\text{ext}} = 0$). The equation then reads

$$\mathcal{H} w = E\, w, \qquad \mathcal{H} = -\frac{\hbar^2}{2\,m}\,\nabla^2 + V. \tag{17.40}$$

Replacing $\mathbf{r}$ with $\mathbf{r} + \mathbf{l}$ is equivalent to add a constant to each component of $\mathbf{r}$, say, $x_i \leftarrow x_i + l_i$, hence the partial derivatives in (17.40) are unaffected. As a

[7] The indices of (16.27) are dropped for simplicity.

consequence, the Hamiltonian operator as a whole has the lattice periodicity, so that its eigenfunctions are Bloch functions. Remembering (17.28), they read

$$w_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r})\, \exp(\mathrm{i}\,\mathbf{k}\cdot\mathbf{r}), \qquad u_{\mathbf{k}}(\mathbf{r}+\mathbf{l}) = u_{\mathbf{k}}(\mathbf{r}), \tag{17.41}$$

with $\mathbf{k}$ belonging to the first Brillouin zone (Sect. 17.5.3). Letting $k^2 = |\mathbf{k}|^2$, (17.41) yields $\nabla^2 w_{\mathbf{k}} = \exp(\mathrm{i}\,\mathbf{k}\cdot\mathbf{r})\,(-k^2+\nabla^2+2\,\mathrm{i}\,\mathbf{k}\cdot\mathrm{grad})\,u_{\mathbf{k}}$ whence, if $E_{\mathbf{k}}$ is the eigenvalue corresponding to $w_{\mathbf{k}}$, the Schrödinger equation (17.40) becomes

$$V\,u_{\mathbf{k}} = \left[E_{\mathbf{k}} + \frac{\hbar^2}{2\,m}\,\left(-k^2+\nabla^2+2\,\mathrm{i}\,\mathbf{k}\cdot\mathrm{grad}\right)\right] u_{\mathbf{k}}. \tag{17.42}$$

As both V and $u_{\mathbf{k}}$ have the periodicity of the lattice, they can be expanded in terms of the translation vectors of the scaled reciprocal lattice $\mathbf{g} = n_1\,2\,\pi\,\mathbf{b}_1 + n_2\,2\,\pi\,\mathbf{b}_2 + n_3\,2\,\pi\,\mathbf{b}_3$:

$$V(\mathbf{r}) = \sum_{\mathbf{g}} V_{\mathbf{g}}\, \exp(\mathrm{i}\,\mathbf{g}\cdot\mathbf{r}), \qquad u_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{g}} s_{\mathbf{k}\mathbf{g}}\, \exp(\mathrm{i}\,\mathbf{g}\cdot\mathbf{r}), \tag{17.43}$$

where $\sum_{\mathbf{g}} = \sum_{n_1}\sum_{n_2}\sum_{n_3}$ and

$$V_{\mathbf{g}} = \frac{1}{\tau_l}\int_{\tau_l} V(\mathbf{r})\, \exp(-\mathrm{i}\,\mathbf{g}\cdot\mathbf{r})\,\mathrm{d}^3 r, \qquad s_{\mathbf{k}\mathbf{g}} = \frac{1}{\tau_l}\int_{\tau_l} u_{\mathbf{k}}(\mathbf{r})\, \exp(-\mathrm{i}\,\mathbf{g}\cdot\mathbf{r})\,\mathrm{d}^3 r. \tag{17.44}$$

Letting $g^2 = |\mathbf{g}|^2$, from the expansion of $u_{\mathbf{k}}$ it follows $(\nabla^2 + 2\,\mathrm{i}\,\mathbf{k}\cdot\mathrm{grad})u_{\mathbf{k}} = -\sum_{\mathbf{g}}(g^2 + 2\,\mathbf{k}\cdot\mathbf{g})\,s_{\mathbf{k}\mathbf{g}}\,\exp(\mathrm{i}\,\mathbf{g}\cdot\mathbf{r})$ whence, using $g^2 + 2\,\mathbf{k}\cdot\mathbf{g} + k^2 = |\mathbf{g}+\mathbf{k}|^2$,

$$V\,u_{\mathbf{k}} = \sum_{\mathbf{g}}\left[E_{\mathbf{k}} - \frac{\hbar^2}{2\,m}\,|\mathbf{g}+\mathbf{k}|^2\right] s_{\mathbf{k}\mathbf{g}}\, \exp(\mathrm{i}\,\mathbf{g}\cdot\mathbf{r}). \tag{17.45}$$

In turn, the left hand side of (17.45) reads

$$V\,u_{\mathbf{k}} = \sum_{\mathbf{g}'}\sum_{\mathbf{g}''} V_{\mathbf{g}'}\, s_{\mathbf{k}\mathbf{g}''}\, \exp[\mathrm{i}\,(\mathbf{g}'+\mathbf{g}'')\cdot\mathbf{r}] = \sum_{\mathbf{g}'}\sum_{\mathbf{g}-\mathbf{g}'} V_{\mathbf{g}'}\, s_{\mathbf{k},\mathbf{g}-\mathbf{g}'}\, \exp(\mathrm{i}\,\mathbf{g}\cdot\mathbf{r}), \tag{17.46}$$

with $\mathbf{g} = \mathbf{g}' + \mathbf{g}''$. Note that the last expression on the right of (17.46) is left unchanged if $\sum_{\mathbf{g}-\mathbf{g}'}$ is replaced with $\sum_{\mathbf{g}}$. In fact, as for each vector $\mathbf{g}'$ the indices n_i of $\mathbf{g}$ span from $-\infty$ to $+\infty$, all ∞^6 combinations of indices of $\mathbf{g}$ and $\mathbf{g}'$ are present in either form of the expansion; using $\sum_{\mathbf{g}}$ instead of $\sum_{\mathbf{g}-\mathbf{g}'}$ merely changes the order of summands. Combining (17.45) with (17.46) then yields

$$\sum_{\mathbf{g}} \exp(\mathrm{i}\,\mathbf{g}\cdot\mathbf{r})\left\{\sum_{\mathbf{g}'} V_{\mathbf{g}'}\, s_{\mathbf{k},\mathbf{g}-\mathbf{g}'} - \left[E_{\mathbf{k}} - \frac{\hbar^2}{2\,m}\,|\mathbf{g}+\mathbf{k}|^2\right] s_{\mathbf{k}\mathbf{g}}\right\} = 0. \tag{17.47}$$

As the factors $\exp(\mathrm{i}\,\mathbf{g}\cdot\mathbf{r})$ are linearly independent from each other for all $\mathbf{r}$, to fulfill (17.47) it is necessary that the term in braces vanishes. To proceed it is useful to associate[8] a single index b to the triad (n_1, n_2, n_3) defining $\mathbf{g}$, and another single index b' to the triad (n'_1, n'_2, n'_3) defining $\mathbf{g}'$. Remembering that at the beginning of this section the assumption of an unlimited lattice has been made, $\mathbf{k}$ must be considered a continuous variable, so that s and E become functions of $\mathbf{k}$ proper. In conclusion, (17.47) transforms into

$$\sum_{b'} s_{b-b'}(\mathbf{k})\, V_{b'} = [E(\mathbf{k}) - T_b(\mathbf{k})]\; s_b(\mathbf{k}), \qquad b = 0, \pm 1, \pm 2, \ldots, \tag{17.48}$$

with $T_b(\mathbf{k})$ the result of the association $b \leftrightarrow (n_1, n_2, n_3)$ in $\hbar^2\,|\mathbf{g}+\mathbf{k}|^2/(2\,m)$. For each $\mathbf{k}$, (17.48) is a linear, homogeneous algebraic system in the infinite unknowns s_b and coefficients $V_{b'}$, $E(\mathbf{k}) - T_b(\mathbf{k})$, with $E(\mathbf{k})$ yet undetermined.

The solution of (17.48) provides an infinite set of eigenvalues $E_1(\mathbf{k})$, $E_2(\mathbf{k})$, $\ldots$, $E_i(\mathbf{k}), \ldots$ associated to the given $\mathbf{k}$. As the latter ranges over the first Brillouin zone, the functions $E_i(\mathbf{k})$ are thought of as branches[9] of a many-valued function. For each branch-index i, the function $E_i(\mathbf{k})$ is called *dispersion relation*, and the set of values spanned by $E_i(\mathbf{k})$ as $\mathbf{k}$ runs over the Brillouin zone is called *energy band* of index i. Being an eigenvalue of a periodic operator, $E_i(\mathbf{k})$ is periodic within the reciprocal, scaled lattice (compare with (17.34)); also, it can be shown that $E_i(\mathbf{k})$ is even with respect to $\mathbf{k}$ (Sect. 17.8.3):

$$E_i(\mathbf{k}+\mathbf{g}) = E_i(\mathbf{k}), \qquad E_i(-\mathbf{k}) = E_i(\mathbf{k}). \tag{17.49}$$

When a finite structure is considered, supplemented with the periodic boundary condition discussed in Sect. 17.5.3, vector $\mathbf{k}$ is discrete. On the other hand, for the derivation of (17.47) it is irrelevant whether $\mathbf{k}$ is continuous or discrete; hence, the analysis carried out in this section still holds for a discrete $\mathbf{k}$, provided the additional relations derived in Sect. 17.5.3, that describe the form of $\mathbf{k}$ and the corresponding densities, are accounted for. It must be remarked that the number N_s of cells along each direction in the direct lattice is typically very large.[10] As a consequence, a change by one unity of n_s in (17.37) is much smaller than the corresponding denominator N_s, so that for all practical purposes $E_i(\mathbf{k})$ is treated as a function of continuous variables when the derivatives with respect to the components of $\mathbf{k}$ enter the calculations.

---

[8] The association $b \leftrightarrow (n_1, n_2, n_3)$ can be accomplished in a one-to-one fashion by, first, distributing the triads into groups having a common value of $d = |n_1| + |n_2| + |n_3|$, then ordering the groups in ascending order of d: for example, $d = 0$ corresponds to $(0,0,0)$, $d = 1$ to $[(0,0,1),(0,1,0),(1,0,0),(0,0,-1),(0,-1,0),(-1,0,0)]$, and so on. As each group is made by construction of a finite number of triads, the latter are numbered within each group using a finite set of values of b; in order to have b ranging from $-\infty$ to $+\infty$, one associates a positive (negative) value of b to the triads in which the number of negative indices is even (odd).

[9] Typically, a graphic representation of $E_i(\mathbf{k})$ is achieved by choosing a crystal direction and drawing the one-dimensional restriction of E_i along such a direction. Examples are given in Sect. 17.6.5.

[10] For instance, in a cube of material with an atomic density of 6.4×10^{27} m^{-3}, the number of atoms per unit length in each direction is 4000 μm^{-1}.

The analysis carried out in this section clarifies the role of $\mathbf{k}$. In fact, for a given band index i, $\mathbf{k}$ labels the energy eigenvalue; for this reason, remembering the discussion about spin carried out in Sect. 15.5, $\mathbf{k}$ and the quantum number associated to spin determine the state of the particle. For fermions, the quantum number associated to spin has two possible values, so that two states with opposite spins are associated to each $\mathbf{k}$ vector. When the periodic boundary conditions are considered, the density of $\mathbf{k}$ vectors in the $\mathbf{k}$ space is given by (17.38), and the combined density of $\mathbf{k}$ vectors in the $\mathbf{r}, \mathbf{k}$ space is given by (17.39). As a consequence, the *density of states* in the $\mathbf{k}$ space and in the $\mathbf{r}, \mathbf{k}$ space are given, respectively, by

$$Q_k = 2\,\frac{N_c}{\tau_g} = \frac{\Omega}{4\,\pi^3}, \qquad Q = 2\,\frac{1}{\Omega}\,\frac{N_c}{\tau_g} = \frac{1}{4\,\pi^3}. \tag{17.50}$$

17.6.1 Wave Packet in a Periodic Potential

From the solution of the Schrödinger equation worked out from (17.48) one reconstructs the periodic part of the Bloch function using the second relation in (17.43). Such a function inherits the band index i, so that the Bloch functions[11] read $w_{i\,\mathbf{k}} = \zeta_{i\,\mathbf{k}}\,\exp(\mathrm{i}\,\mathbf{k}\cdot\mathbf{r})$; they form a complete set so that, letting $\omega_{i\,\mathbf{k}} = \omega_i(\mathbf{k}) = E_i(\mathbf{k})/\hbar$, the expansion of the wave function $\psi(\mathbf{r}, t)$ in terms of the eigenfunctions of the periodic Hamiltonian operator (17.40) reads

$$\psi(\mathbf{r}, t) = \sum_{i\,\mathbf{k}} c_{i\,\mathbf{k}}\, w_{i\,\mathbf{k}}(\mathbf{r})\,\exp(-\mathrm{i}\,\omega_{i\,\mathbf{k}}\,t) = \sum_{i\,\mathbf{k}} c_{i\,\mathbf{k}}\,\zeta_{i\,\mathbf{k}}(\mathbf{r})\,\exp[\mathrm{i}\,(\mathbf{k}\cdot\mathbf{r} - \omega_{i\,\mathbf{k}}\,t)], \tag{17.51}$$

with $c_{i\,\mathbf{k}} = \langle w_{i\,\mathbf{k}}|\psi\rangle_{t=0}$ a set of constants. The expansion (17.51) bears a strong similarity with that of the wave packet describing a free particle (compare with (9.1)), the only difference between the two expansions being the periodic factor $\zeta_{i\,\mathbf{k}}(\mathbf{r})$. The similarity suggests that an approximate expression of the wave packet is achieved by following the same reasoning as in Sect. 9.6, namely, by expanding $\omega_i(\mathbf{k})$ around the average value $\mathbf{k}_0$ of the wave vector and retaining the first-order term of the expansion:[12]

$$\omega_i(\mathbf{k}) \simeq \omega_i(\mathbf{k}_0) + \mathbf{u}_i(\mathbf{k}_0)\cdot(\mathbf{k} - \mathbf{k}_0), \qquad \mathbf{u}_i(\mathbf{k}_0) = \left(\mathrm{grad}_{\mathbf{k}}\,\omega_i\right)_0, \tag{17.52}$$

with $\mathbf{u}_i$ the group velocity of the ith band. The approximation holds as long as $|R_i|\,t \ll 2\,\pi$, where R_i is the rest of the expansion. Letting $\omega_{i\,0} = \omega_i(\mathbf{k}_0)$, $\zeta_{i\,0} = \zeta_i(\mathbf{k}_0)$, and $\Phi_{i\,0} = \mathbf{k}_0\cdot\mathbf{r} - \omega_{i\,0}\,t$, the approximate expression of ψ reads

$$\psi(\mathbf{r}, t) \simeq \sum_{i\,\mathbf{k}} c_{i\,\mathbf{k}}\,\zeta_{i\,0}\,\exp(\mathrm{i}\,\Phi_{i\,0})\,\exp[\mathrm{i}\,(\mathbf{r} - \mathbf{u}_i\,t)\cdot(\mathbf{k} - \mathbf{k}_0)]\,. \tag{17.53}$$

[11] Here the periodic part of $w_{i\,\mathbf{k}}$ is indicated with $\zeta_{i\,\mathbf{k}}$ to avoid confusion with the group velocity.

[12] In the case of a free particle (Sect. 9.6) the approximation neglects only the second order because $\omega(\mathbf{k})$ has a quadratic dependence on the components of $\mathbf{k}$. Here, instead, the expansion has in general all terms due to the more complicate form of $\omega_i(\mathbf{k})$, so the neglected rest R_i contains infinite terms.

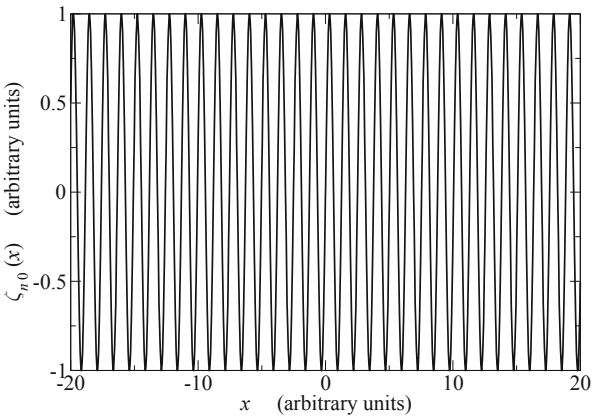

Fig. 17.7 A one-dimensional example of the periodic factor $\zeta_{n\,0}$ of (17.56)

The envelope function is now defined as in (9.26), the difference being that a sum is used here instead of an integral:

$$A\,(\mathbf{r}-\mathbf{u}_i\,t\,;\mathbf{k}_0)=\sum_{\mathbf{k}} c_{i\,\mathbf{k}}\exp\left[\mathrm{i}\;(\mathbf{r}-\mathbf{u}_i\,t)\cdot(\mathbf{k}-\mathbf{k}_0)\right], \tag{17.54}$$

so that

$$\psi(\mathbf{r},t)\simeq\sum_i \zeta_{i\,0}\;\exp\left(\mathrm{i}\,\Phi_{i\,0}\right)A\,(\mathbf{r}-\mathbf{u}_i\,t\,;\mathbf{k}_0)\,. \tag{17.55}$$

As a further approximation one considers the fact that the number of $\mathbf{k}$ vectors of the first Brillouin zone is in general very large, because it equals the number N_c of direct-lattice cells. It follows that, although the set of eigenfunctions belonging to a single branch is not complete, such a set is still able to provide an acceptable description of the wave packet. In this case one fixes the branch index, say, $i=n$, so that $\psi(\mathbf{r},t)\simeq\zeta_{n\,0}\;\exp\left(\mathrm{i}\,\Phi_{n\,0}\right)A(\mathbf{r}-\mathbf{u}_n\,t\,;\mathbf{k}_0)$. It follows

$$|\psi(\mathbf{r},t)|^2\simeq|\zeta_{n\,0}(\mathbf{r})|^2\,|A\,(\mathbf{r}-\mathbf{u}_n\,t\,;\mathbf{k}_0)\,|^2. \tag{17.56}$$

In (17.56), the periodic factor $|\zeta_{n\,0}(\mathbf{r})|^2$ is a rapidly-oscillating term whose period is of the order of the lattice constant; such a term does not provide any information about the particle's localization. This information, in fact, is carried by the envelope function, like in the case of a free particle outlined in Sect. 9.6. A one-dimensional example about how (17.56) is built up is given in Figs. 17.7, 17.8, 17.9, and 17.10.

17.6.2 Parabolic-Band Approximation

The dispersion relation $E_n(\mathbf{k})$ obtained from the solution of the Schrödinger equation (17.40) in a periodic lattice fulfills the periodicity condition given by the

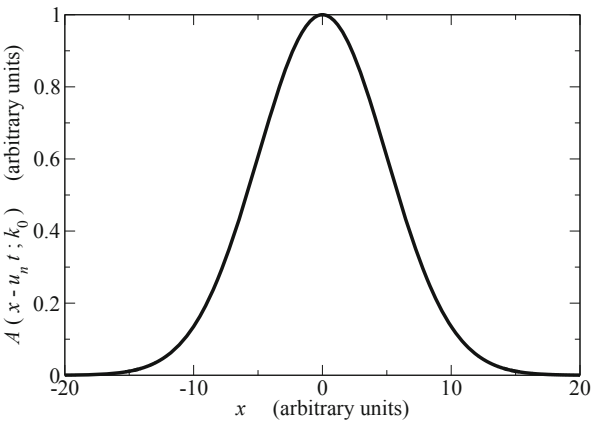

Fig. 17.8 A one-dimensional example of the envelope function $A(\mathbf{r} - \mathbf{u}_n\, t\,; \mathbf{k}_0)$ of (17.56)

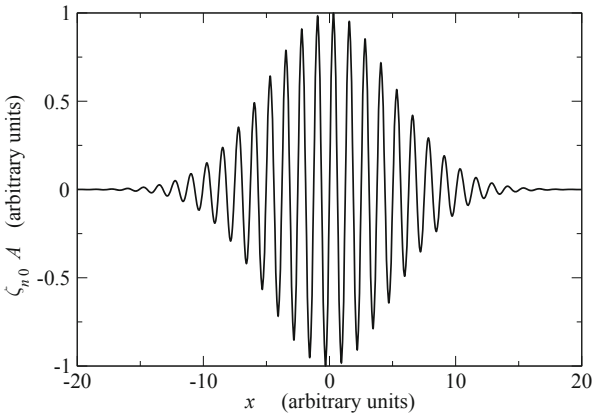

Fig. 17.9 Product of the two functions shown in Figs. 17.7 and 17.8

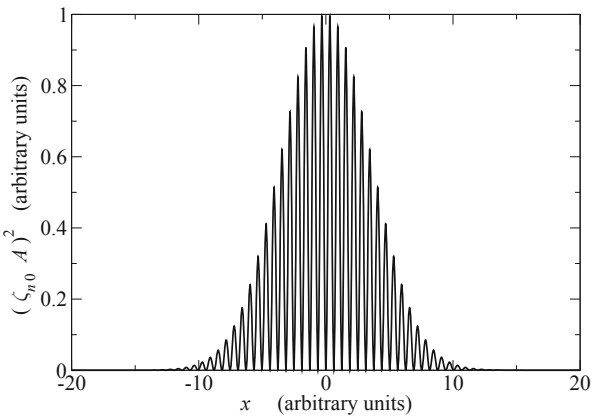

Fig. 17.10 The function of Fig. 17.9 squared

first expression in (17.49). As a consequence, $E_n(\mathbf{k})$ has necessarily a number of extremum points within the first Brillouin zone or at the boundary of it.[13]

In view of further developments of the theory it is useful to investigate the form of $E_n(\mathbf{k})$ in the vicinity of such extremum points. To this purpose, the absolute minima are considered first; for a given branch index n assume that the number of such minima is M_C, and let $\mathbf{k}_a$ be the value of $\mathbf{k}$ at the ath minimum, $a = 1, \ldots, M_C$, with $E_C = E_n(\mathbf{k}_a)$. At $\mathbf{k} = \mathbf{k}_a$ the Hessian matrix of $E_n(\mathbf{k})$ is symmetric and positive definite, hence it can be diagonalized with positive real eigenvalues. In other terms, the reference in the $\mathbf{k}$ space can be chosen in such a way as to make the Hessian matrix of $E_n(\mathbf{k})$ diagonal; using such a reference, the second-order expansion of $E_n(\mathbf{k})$ around $\mathbf{k}_a$ reads

$$E_n(\mathbf{k}) \simeq E_C + \frac{1}{2} \sum_{i=1}^{3} \left(\frac{\partial^2 E_n}{\partial k_i^2} \right)_a (k_i - k_{ia})^2 \geq E_a, \quad a = 1, \ldots, M_C. \tag{17.57}$$

The first derivatives are missing from (17.57) because the expansion is carried out at an extremum. The coefficients $(\partial^2 E_n / \partial k_i^2)_a$ are in general different from each other, so that the sum in (17.57) may be thought of as a positive-definite quadratic form generated by a 3×3 diagonal matrix. Noting the units of the matrix entries one defines the *inverse, effective-mass tensor* of the ath minimum as

$$(\hat{m}_a)^{-1} = \begin{bmatrix} 1/m_{1a} & 0 & 0 \\ 0 & 1/m_{2a} & 0 \\ 0 & 0 & 1/m_{3a} \end{bmatrix}, \qquad \frac{1}{m_{ia}} = \frac{1}{\hbar^2} \left(\frac{\partial^2 E_n}{\partial k_i^2} \right)_a > 0 \tag{17.58}$$

so that, using the notation $E_{ne}(\mathbf{k}) = E_n(\mathbf{k}) - E_C \geq 0$, $\delta k_i = k_i - k_{ia}$, (17.57) takes the form

$$E_{ne} = \sum_{i=1}^{3} \frac{\hbar^2}{2\, m_{ia}} (k_i - k_{ia})^2 = \frac{1}{2}\, \hbar\, \delta\mathbf{k} \cdot (\hat{m}_a)^{-1}\, \hbar\, \delta\mathbf{k}. \tag{17.59}$$

Being the inverse, effective-mass tensor diagonal, the *effective-mass tensor* $\hat{m}_a$ is given by

$$\hat{m}_a = \begin{bmatrix} m_{1a} & 0 & 0 \\ 0 & m_{2a} & 0 \\ 0 & 0 & m_{3a} \end{bmatrix}. \tag{17.60}$$

The approximation shown above, that consists in replacing the dispersion relation with its second-order expansion near an extremum, is called *parabolic-band approximation*. The group velocity to be associated with a $\mathbf{k}$ vector in the vicinity of a

[13] As mentioned in Sect. 17.6, $E_n(\mathbf{k})$ is considered as a function of a continuous vector variable $\mathbf{k}$ even when the periodic boundary conditions are assumed.

minimum is found by applying to (17.59) the second relation in (17.52):

$$\mathbf{u}_n(\mathbf{k}) = \frac{1}{\hbar}\,\mathrm{grad}_{\mathbf{k}}\,E_n(\mathbf{k}) = \frac{1}{\hbar}\,\mathrm{grad}_{\mathbf{k}}\,E_{ne}(\mathbf{k}) = (\hat{m}_a)^{-1}\,\hbar\,\delta\mathbf{k}. \tag{17.61}$$

The calculation in the vicinity of an absolute maximum is similar.[14] Assume that the number of maxima in the nth branch of the dispersion relation is M_V, and let $\mathbf{k}_a$ be the value of $\mathbf{k}$ at the ath maximum, $a = 1, \ldots, M_V$, with $E_V = E_n(\mathbf{k}_a)$. The second-order expansion of $E_n(\mathbf{k})$ around $\mathbf{k}_a$ reads

$$E_n(\mathbf{k}) \simeq E_V + \frac{1}{2}\sum_{i=1}^{3}\left(\frac{\partial^2 E_n}{\partial k_i^2}\right)_a (k_i - k_{ia})^2 \leq E_a, \quad a = 1, \ldots, M_V, \tag{17.62}$$

where the Hessian matrix is negative definite. For this reason, the inverse, effective-mass tensor at the ath maximum is defined as

$$(\hat{m}_a)^{-1} = \begin{bmatrix} 1/m_{1a} & 0 & 0 \\ 0 & 1/m_{2a} & 0 \\ 0 & 0 & 1/m_{3a} \end{bmatrix}, \qquad \frac{1}{m_{ia}} = -\frac{1}{\hbar^2}\left(\frac{\partial^2 E_n}{\partial k_i^2}\right)_a > 0 \tag{17.63}$$

so that, using the notation $E_{nh}(\mathbf{k}) = E_V - E_n(\mathbf{k}) \geq 0$, (17.57) takes the form

$$E_{nh} = \sum_{i=1}^{3}\frac{\hbar^2}{2\,m_{ia}}\,(k_i - k_{ia})^2 = \frac{1}{2}\,\hbar\,\delta\mathbf{k}\cdot(\hat{m}_a)^{-1}\,\hbar\,\delta\mathbf{k}. \tag{17.64}$$

The group velocity to be associated with a $\mathbf{k}$ vector in the vicinity of a maximum reads

$$\mathbf{u}_n(\mathbf{k}) = \frac{1}{\hbar}\,\mathrm{grad}_{\mathbf{k}}\,E_n(\mathbf{k}) = -\frac{1}{\hbar}\,\mathrm{grad}_{\mathbf{k}}\,E_{nh}(\mathbf{k}) = -(\hat{m}_a)^{-1}\,\hbar\,\delta\mathbf{k}. \tag{17.65}$$

It is important to note that the expressions of the parabolic-band approximation given in this section have been worked out in a specific reference of the $\mathbf{k}$ space, namely, the reference where the Hessian matrix is diagonal. In so doing, the reference of the direct space $\mathbf{r}$ has been fixed as well, because the two references are reciprocal to each other (Sect. 17.3). In other terms, when diagonal expressions like (17.59) or (17.64) are used in a dynamical calculation, the reference in the $\mathbf{r}$ space can not be chosen arbitrarily.

[14] The parabolic-band approximation is not necessarily limited to absolute minima or absolute maxima; here it is worked out with reference to such cases because they are the most interesting ones. However, it applies as well to relative minima and relative maxima. The different values of the inverse, effective-mass tensor's entries between an absolute and a relative minimum of a branch in GaAs give rise to interesting physical effects (Sect. 17.6.6).

17.6.3 Density of States in the Parabolic-Band Approximation

Calculations related to many-particle systems often involve the density of states in energy (e.g., Sects. 15.8.1, 15.8.2). The calculation of this quantity is relatively simple for the dispersion relation $E_n(\mathbf{k})$ in the parabolic-band approximation, because the dispersion relation is quadratic in the components of $\mathbf{k}$ and, in turn, the density of the $\mathbf{k}$ vectors is constant. In fact, it is found from (B.34) that in the three-dimensional case a quadratic expression $A = u^2 + v^2 + w^2$ yields a density of states equal to $A^{1/2}$. It is then sufficient to reduce (17.59) and (17.64) to the quadratic expression above. Taking (17.59) by way of example, one disposes with the multiplicative factors and the shift in the origin by applying the *Herring–Vogt transformation*

$$\eta_i = \frac{\hbar}{\sqrt{2\, m_{ia}}} (k_i - k_{ia}), \tag{17.66}$$

to find

$$E_{ne} = \sum_{i=1}^{3} \eta_i^2 = \eta^2, \qquad \mathrm{d}k_i = \frac{\sqrt{2\, m_{ia}}}{\hbar}\, \mathrm{d}\eta_i, \tag{17.67}$$

with $\eta > 0$, and

$$\mathrm{d}^3k = \mathrm{d}k_1\, \mathrm{d}k_2\, \mathrm{d}k_3 = 2\frac{\sqrt{2}}{\hbar^3}\, m_{ea}^{3/2}\, \mathrm{d}^3\eta, \qquad m_{ea} = (m_{1a}\, m_{2a}\, m_{3a})^{1/3}. \tag{17.68}$$

Turning to spherical coordinates $\eta_1 = \eta \sin\vartheta \cos\varphi$, $\eta_2 = \eta \sin\vartheta \sin\varphi$, $\eta_3 = \eta \cos\vartheta$ yields $\mathrm{d}^3\eta = \eta^2 \mathrm{d}\eta \sin\vartheta\, \mathrm{d}\vartheta\, \mathrm{d}\varphi$ (Sect. B.1), where the product $\eta^2\, \mathrm{d}\eta$ is found by combining the relations $2\,\eta\, \mathrm{d}\eta = \mathrm{d}E_{ne}$ and $\mathrm{d}\eta = \mathrm{d}E_{ne}/(2\sqrt{E_{ne}})$:

$$\eta^2\, \mathrm{d}\eta = \frac{1}{2}\sqrt{E_{ne}}\, \mathrm{d}E_{ne}. \tag{17.69}$$

The number of states belonging to the elementary volume d^3k is $\mathrm{d}N = Q_k\, \mathrm{d}^3k$, with Q_k the density of states in the $\mathbf{k}$ space given by the first expression in (17.50). If the elementary volume is centered on a $\mathbf{k}$ vector close to the ath minimum of $E_n(\mathbf{k})$, so that the parabolic-band approximation holds, one has

$$\mathrm{d}N_a = Q_k\, \mathrm{d}^3k = \frac{\Omega}{4\,\pi^3}\, 2\, \frac{\sqrt{2}}{\hbar^3}\, m_{ea}^{3/2}\, \frac{1}{2}\sqrt{E_{ne}}\, \mathrm{d}E_{ne}\, \sin\vartheta\, \mathrm{d}\vartheta\, \mathrm{d}\varphi. \tag{17.70}$$

The integral over the angles yields 4π, whence

$$\int_{\varphi=0}^{2\pi} \int_{\vartheta=0}^{\pi} \mathrm{d}N_a = \Omega\, \frac{\sqrt{2}}{\pi^2\, \hbar^3}\, m_{ea}^{3/2}\, \sqrt{E_{ne}}\, \mathrm{d}E_{ea} = g_a(E_{ne})\, \mathrm{d}E_{ne} \tag{17.71}$$

where, by construction, $g_a(E_{ne})$ is the density of states in energy around the ath minimum. Adding g_a over the M_C absolute minima yields the total density of states in energy,

$$g(E_{ne}) = \sum_{a=1}^{M_C} g_a = \Omega \frac{\sqrt{2}}{\pi^2 \hbar^3} M_C \, m_e^{3/2} \sqrt{E_{ne}}, \qquad m_e = \left(\frac{1}{M_C} \sum_{a=1}^{M_C} m_{ea}^{3/2} \right)^{2/3}, \tag{17.72}$$

with m_e the *average effective mass* of the absolute minima. The *combined density of states* in the energy and **r** spaces then reads

$$\gamma(E_{ne}) = \frac{g(E_{ne})}{\Omega} = \frac{\sqrt{2}}{\pi^2 \hbar^3} M_C \, m_e^{3/2} \sqrt{E_{ne}}. \tag{17.73}$$

Note that, apart from the different symbol used to indicate the volume in the **r** space, and the replacement of m with $M_C \, m_e^{3/2}$, the relations (17.72) and (17.73) are identical, respectively, to (15.64) and (15.65), expressing the density of states and combined density of states in a box.

The calculation of the density of states in energy in the vicinity of the M_V absolute maxima is identical to the above, and yields

$$g(E_{nh}) = \sum_{a=1}^{M_V} g_a = \Omega \frac{\sqrt{2}}{\pi^2 \hbar^3} M_V \, m_h^{3/2} \sqrt{E_{nh}}, \quad m_h = \left(\frac{1}{M_V} \sum_{a=1}^{M_V} m_{ha}^{3/2} \right)^{2/3}, \tag{17.74}$$

where m_h is the average effective mass of the absolute maxima. In turn it is $m_{ha} = (m_{1a} \, m_{2a} \, m_{3a})^{1/3}$, with m_{ia} given by the second relation in (17.63).

17.6.4 Crystals of Si, Ge, and GaAs

Among semiconductors, silicon (Si), germanium (Ge), and gallium arsenide (GaAs) are very important for the electronic industry. This section is devoted to illustrating some properties of their crystal and energy-band structures. The crystals of silicon and germanium are of the face-centered, cubic type; the reciprocal lattices have the body-centered, cubic structure. The *lattice constants*, that is, the physical sizes of the unit cell, are the same in the [100], [010], and [001] directions (Sect. 17.8.1). Their values at $T = 300$ K are given in Table 17.1 [80]. The crystals of the materials under consideration are formed by elementary blocks like that shown in Fig. 17.11. Each atom has four electrons in the external shell, so that it can form four chemical bonds with other identical atoms; the latter place themselves symmetrically in space, to build up the tetrahedral structure shown in the figure. In this structure, which is of the body-centered cubic type with a side equal to one half the lattice constant a, the chemical bonds of the central atom are saturated, whereas the atoms placed at

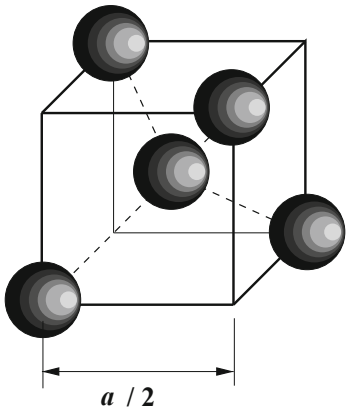

Fig. 17.11 Tetrahedral organization of the elementary, body-centered cubic block of silicon or germanium. The side of the cube is one half the lattice constant a

the vertices still have three bonds unsaturated; as a consequence, they may behave as centers of new tetrahedral structures identical to the original one.

An example of this is given in Fig. 17.12: the top half of the figure shows two replicas of the elementary block of Fig. 17.11 sharing an atom belonging to an upper corner, while the bottom half of the figure shows again two replicas, this time sharing an atom belonging to a lower corner. The atoms drawn in white do not belong to any of the elementary blocks considered in the figure, and serve the purpose of demonstrating how the rest of the crystal is connected to them. Note that the structure in the bottom half of Fig. 17.12 is identical to that of the top half, the difference being simply that one structure is rotated by 90° with respect to the other on a vertical axis. The construction is now completed by bringing the two halves together, as shown in Fig. 17.13; this provides the diamond structure mentioned in Sect. 17.2. Such a structure is of the face-centered, cubic type, with an additional atom at the center of each tetrahedral block.

The minimum distance d among the atoms (*interatomic distance*) is the distance from the atom in the center of the tetrahedral elementary block to any of the atoms at its vertices; its relation with the lattice constant is easily found to be

$$d = \frac{\sqrt{3}}{4} a. \tag{17.75}$$

The description is similar for gallium arsenide [80], and for a number of semiconductors of the III-V type, whose crystal constants are listed in Table 17.2.

17.6.5 Band Structure of Si, Ge, and GaAs

Coming now to the description of the band structure, it is important to focus on the bands that are able to contribute to the electric conduction of the material. In fact, considering the aim of manufacturing electronic devices out of these materials, the

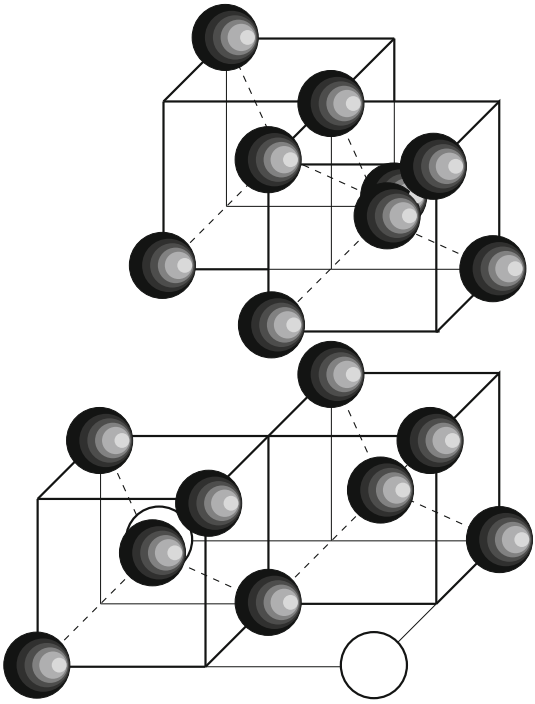

Fig. 17.12 Diamond structure. The *top* and *bottom* halves are shown separately

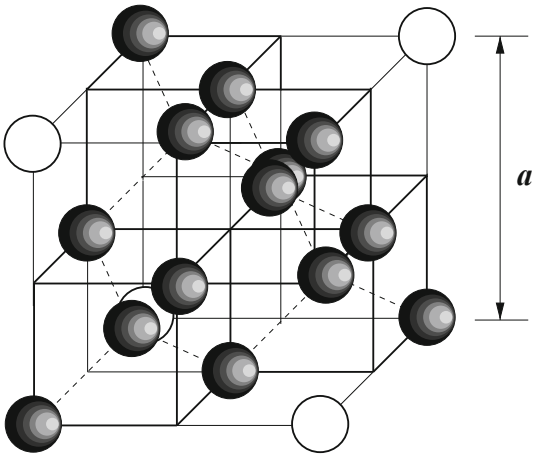

Fig. 17.13 Diamond structure obtained by joining together the *top* and *bottom* halves shown separately in Fig. 17.12

bands that do not contribute to the electric current are not relevant. It is intuitive that a band with no electrons, that is, whose states have a zero probability of being occupied, is not able to provide any conduction; it is less intuitive (in fact, this is demonstrated in Sect. 19.3) that a band whose states are fully occupied does not provide any conduction either. It follows that the only bands of interest are those

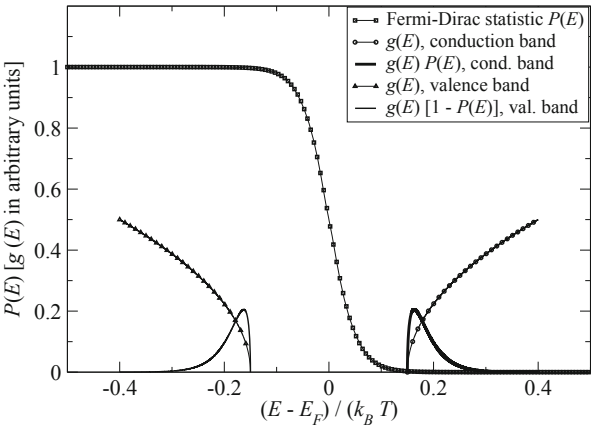

Fig. 17.14 Calculation of the particles' population in the conduction and valence bands of a semiconductor. To make them more visible, the products $g(E)\,P(E)$ and $g(E)\,[1 - P(E)]$ have been amplified with respect to $g(E)$ alone. The gap's extension is arbitrary and does not refer to any specific material

where only a fraction of the electronic states are occupied. Although a discussion involving the electric current must necessarily refer to a non-equilibrium condition, it is easier to base the reasoning upon the equilibrium condition at some temperature T; in fact, in this case the occupation probability of the electronic states is given by the Fermi–Dirac statistics (15.49). As a consequence, the number of electrons belonging to a band whose energy values range, say, from E_a to E_b, is given by the first relation in (15.48) with $\alpha + \beta\,E = (E - E_F)/(k_B\,T)$, namely,

$$N_{ab} = \int_{E_a}^{E_b} \frac{g(E)}{\exp\left[(E - E_F)/(k_B\,T)\right] + 1}\,\mathrm{d}E. \tag{17.76}$$

As mentioned in Sect. 17.6 each branch of the dispersion relation $E_i(\mathbf{k})$ spans an energy band. In many cases the bands are disjoint from each other, namely, energy intervals exist that contain no eigenvalue of the Schrödinger equation (17.40). Such intervals are called *forbidden bands* or *gaps*. In the equilibrium condition the energy of an electron can never belong to a gap, no matter what the value of the occupation probability is, because the density of states is zero there. Also, at a given temperature the position of the Fermi level E_F is either within a band (edges included), or within a gap; the latter case, typical of semiconductors, is illustrated with the aid of Fig. 17.14, where it is assumed (using the units of $(E - E_F)/(k_B\,T)$) that a gap exists between the energies E_V, E_C such that $(E_V - E_F)/(k_B\,T) = -0.15$ and $(E_C - E_F)/(k_B\,T) = +0.15$. In other terms, E_V is the upper energy edge of a band, and E_C the lower energy edge of the next band. These assumptions also imply that the Fermi level coincides with the gap's midpoint. As will become apparent below, the two bands that are separated by the Fermi level are especially important; for this reason they are given specific names: the band whose absolute maximum is E_V is called *valence band*, that whose absolute minimum is E_C is called *conduction band*.

As shown in the figure, the case is considered (typical of Si, Ge, and GaAs) where the gap's width contains the main variation of the Fermi–Dirac statistics;[15] as a consequence, the occupation probability becomes vanishingly small as the difference $E - E_C$ becomes larger, so that only the energy states near the absolute minimum E_C have a non-vanishing probability of being occupied. Thank to this reasoning, to the purpose of calculating (17.76) one can replace the density of states $g(E)$ with the simplified expression (17.72) deduced from the parabolic-band approximation; such an expression, $g(E) \propto \sqrt{E - E_C}$, is shown in Fig. 17.14 in arbitrary units, along with the $g(E)\,P(E)$ product (thick line), that represents the integrand of (17.76) with reference to the conduction band. To make it more visible, the $g(E)\,P(E)$ product is drawn in a scale amplified by 10^3 with respect to that of $g(E)$ alone. The number of electrons belonging to the conduction band is proportional to the area subtended by the $g(E)\,P(E)$ curve.

Coming now to the valence band, the probability $1 - P(E)$ that a state at energy E is empty becomes vanishingly small as the difference $E_V - E$ becomes larger, so that only the energy states near the absolute maximum E_V have a non-vanishing probability of being empty. Empty states are also called *holes*. The number of holes is given by an integral similar to (17.76), where $P(E)$ is replaced with $1 - P(E)$. This calculation is made easier by observing that, due to the form of the Fermi–Dirac statistics, it is

$$1 - \frac{1}{\exp\left[(E - E_F)/(k_B\,T)\right] + 1} = \frac{1}{\exp\left[(E_F - E)/(k_B\,T)\right] + 1}. \tag{17.77}$$

Also in this case one can use for the density of states the parabolic-band approximation; such an expression, $g(E) \propto \sqrt{E_V - E}$, is shown in Fig. 17.14 in arbitrary units, along with the $g(E)\,[1 - P(E)]$ product (thin line). As before, the product is drawn in a scale amplified by 10^3 with respect to that of $g(E)$ alone. The number of holes belonging to the valence band is proportional to the area subtended by the $g(E)\,[1 - P(E)]$ curve.

Thanks to the spatial uniformity of the crystal, the concentration of the electrons in the conduction band is obtained by dividing their number by the crystal volume Ω or, equivalently, by replacing the density of states in energy g with the combined density of states in energy and volume γ given by (17.73). A similar reasoning holds for holes. The explicit expressions of the concentrations are given in Sect. 18.3. Here it is important to remark that the perfect symmetry of the curves $g(E)\,P(E)$ and $g(E)\,[1 - P(E)]$ in Fig. 17.14 is due to the simplifying assumptions that E_F coincides with the gap's midpoint and that $M_V\,m_h^{3/2} = M_C\,m_e^{3/2}$ (compare with (17.73) and (17.74)). Neither hypothesis is actually true, so that in real cases the two curves are not symmetric; however, as shown in Sect. 18.3, the areas subtended by them are nevertheless equal to each other.

[15] The extension of the energy region where the main variation of the Fermi statistics occurs is estimated in Prob. 15.1.

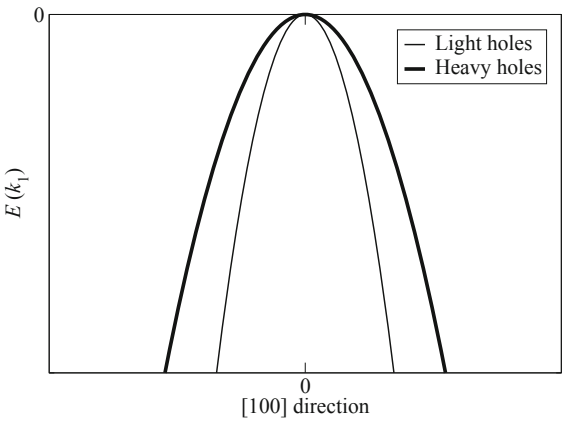

Fig. 17.15 Schematic view of the two branches of the valence band of Si, Ge, or GaAs in the [100] direction

17.6.5.1 Valence Band

The valence band of Si, Ge, and GaAs is made of two branches of $E(\mathbf{k})$, having the same absolute maximum E_V at $\mathbf{k} = 0$ (so that $M_V = 2$), but different curvatures. They are shown in Fig. 17.15, where the horizontal axis coincides with the [100] direction in the $\mathbf{k}$ space, corresponding to the scalar variable k_1. As a consequence, the origin of the horizontal axis coincides with the Γ point (Sect. 17.4); the axis intersects the boundary of the first Brillouin zone at the X points (not shown in the figure). The origin of the vertical axis coincides with E_V. The two branches are not spherically symmetric; in fact, letting $E_V = 0$, the dependence of each of them on the spherical coordinates k, ϑ, φ has the form [54, Sect. 8.7]

$$-\frac{\alpha}{2} k^2 \left[1 \pm j(\vartheta, \varphi)\right], \qquad \alpha > 0, \tag{17.78}$$

called *warped*. In the parabolic-band approximation the angular part j is neglected with respect to unity, and the two branches become spherically symmetric around $\mathbf{k} = 0$; still with $E_V = 0$, the dependence on k_1 of each branch has the form $E = -\alpha k_1^2/2$, where the constant α is smaller in the upper branch (indicated by the thick line in Fig. 17.15), and larger in the lower one. As a consequence, the corresponding component of the effective-mass tensor (17.63), that reads in this case $m_1 = \hbar^2/\alpha$, is larger in the upper branch and smaller in the lower one. For this reason, the holes associated to the energy states of the upper branch are called *heavy holes*, those associated to the lower branch are called *light holes*.

The analysis is identical in the other two directions [010] and [001] so that, for each branch of the valence band, the diagonal entries of the effective-mass tensor are equal to each other. Such tensors then read $m_{hh}\,\mathcal{I}$, $m_{hl}\,\mathcal{I}$, with $\mathcal{I}$ the identity tensor; the first index of the scalar effective mass stands for "hole", while the second one stands for "heavy" or "light". The second-order expansions around E_V take

Table 17.3 Normalized effective masses of the valence band of Si, Ge, and GasAs

Material	$m_{hh}(T_a)/m_0$	$m_{hl}(T_a)/m_0$
Si	0.5	0.16
Ge	0.3	0.04
GaAs	0.5	0.12

respectively the form[16]

$$E_V - E_h(\mathbf{k}) = \frac{\hbar^2}{2\,m_{hh}} \sum_{i=1}^{3} k_i^2, \qquad E_V - E_l(\mathbf{k}) = \frac{\hbar^2}{2\,m_{hl}} \sum_{i=1}^{3} k_i^2. \tag{17.79}$$

Due to (17.79), the constant-energy surfaces $E_V - E_h(\mathbf{k}) = \text{const}$ and $E_V - E_l(\mathbf{k}) = \text{const}$ are spheres, whose radius squared is $2\,m_{hh}\,(E_V - E_h)/\hbar^2$ and $2\,m_{hl}\,(E_V - E_l)/\hbar^2$, respectively. The values of m_{hh} and m_{hl} at room temperature T_a are listed in Table 17.3 [103, Sect. 2-3]; they are normalized to the rest mass of the free electron, $m_0 \simeq 9.11 \times 10^{-31}$ kg. The effective masses depend in general on temperature because a change in the latter modifies the lattice constants: as a consequence, the characteristic vectors of the reciprocal lattice change as well, this deforming the dispersion relation $E(\mathbf{k})$; on the other hand, the variation of the effective masses with temperature is weak, so it is typically neglected.[17]

17.6.5.2 Conduction Band

The conduction band of Si, Ge, and GaAs has only one branch. However, the absolute minima (also called *valleys*) are placed differently. In GaAs there is only one absolute minimum at $\mathbf{k} = 0$, with spherical symmetry. In the parabolic-band approximation, the constant-energy surface is given by

$$E(\mathbf{k}) - E_C = \frac{\hbar^2}{2\,m_e} \sum_{i=1}^{3} k_i^2, \tag{17.80}$$

namely, a sphere whose radius squared is $2\,m_e\,(E - E_C)/\hbar^2$. The band exhibits also secondary minima at $E_C + \Delta E$, with $\Delta E \simeq 0.36$ eV (Fig. 17.16).

The conduction band of Si has six absolute minima ($M_C = 6$), grouped into three pairs. The latter belong to the [100], [010], and [001] directions, respectively, and are symmetrically placed with respect to the Γ point $\mathbf{k} = 0$. Their coordinates are

$$[100] : (\pm k_m, 0, 0), \qquad [010] : (0, \pm k_m, 0), \qquad [001] : (0, 0, \pm k_m), \tag{17.81}$$

where $k_m \simeq 0.85\,k_B > 0$, with k_B the distance between the Γ and X points (Fig. 17.17).

[16] From now on the band index n introduced in (17.56) is omitted from the notation.

[17] In contrast, the temperature dependence of the energy gap, due to the deformation of the dispersion relation, can not be neglected because of its strong effect on the carrier concentration (Sect. 18.3).

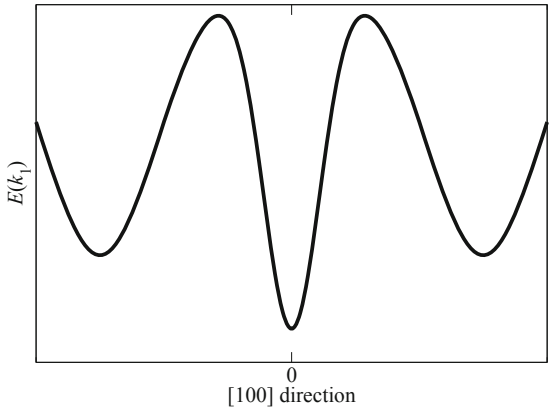

Fig. 17.16 Schematic view of the conduction band of GaAs in the [100] direction

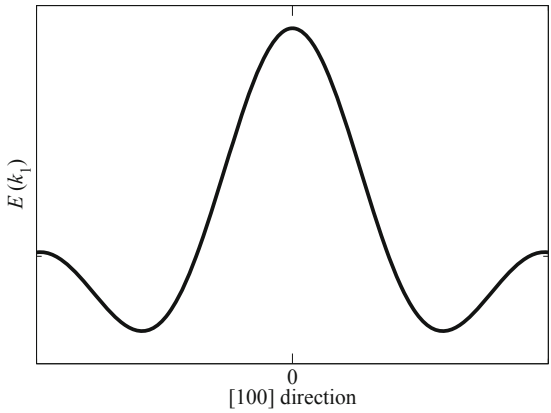

Fig. 17.17 Schematic view of the conduction band of Si in the [100] direction

In the parabolic-band approximation, the surfaces at constant energy of the conduction band of Si are ellipsoids of revolution about the [100], [010], or [001] axes. Their expressions are

$$[100]: \qquad E_{e1} = E(\mathbf{k}) - E_C = \frac{\hbar^2}{2}\left[\frac{(k_1 - k_m)^2}{m_l} + \frac{k_2^2}{m_t} + \frac{k_3^2}{m_t}\right], \tag{17.82}$$

$$[010]: \qquad E_{e2} = E(\mathbf{k}) - E_C = \frac{\hbar^2}{2}\left[\frac{k_1^2}{m_t} + \frac{(k_2 - k_m)^2}{m_l} + \frac{k_3^2}{m_t}\right], \tag{17.83}$$

$$[001]: \qquad E_{e3} = E(\mathbf{k}) - E_C = \frac{\hbar^2}{2}\left[\frac{k_1^2}{m_t} + \frac{k_2^2}{m_t} + \frac{(k_3 - k_m)^2}{m_l}\right]. \tag{17.84}$$

Similarly, E_{e4}, E_{e5}, E_{e6} are derived from E_{e1}, E_{e2}, E_{e3}, respectively, by letting $k_m \leftarrow -k_m$. The effective masses m_l and m_t are called *longitudinal* and *transverse* mass, respectively.

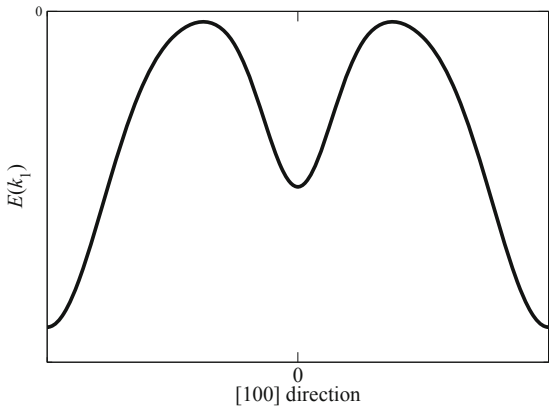

Fig. 17.18 Schematic view of the conduction band of Ge in the [100] direction.

Table 17.4 Normalized effective masses of the conduction band of Si, Ge, and GasAs

Material	$m_l(T_a)/m_0$	$m_t(T_a)/m_0$
Si	0.97	0.19
Ge	1.6	0.082
GaAs[a]	0.068	0.068

[a]The effective masses of GaAs are equal to each other due to the band's isotropy

The conduction band of Ge has eight absolute minima, grouped into four pairs. The pairs belong to the four {111} directions and are placed at the boundary of the first Brillouin zone (Fig. 17.18); thus, only four absolute minima must be accounted for ($M_C = 4$). In the parabolic-band approximation, the surfaces at constant energy of the conduction band of Ge are ellipsoids of revolution about the corresponding axis; like in silicon, for each ellipsoid the longitudinal mass corresponds to the direction of the axis itself, while the transverse masses correspond to the directions normal to it. The values of m_l and m_t at room temperature T_a, normalized to the rest mass of the free electron, are listed in Table 17.4 [103, Sect. 2-3].

17.6.6 Further Comments About the Band Structure

As better detailed in Sects. 19.5.2 and 19.5.3, among the coefficients of the equations describing the transport phenomena in a semiconductor are the electron and hole *mobilities*, that enter the relation between current density and electric field in a uniform material. For the conduction band of the semiconductors considered here, and in the parabolic-band approximation, the electron mobility μ_n turns out to be proportional to $1/m_n = (2/m_t + 1/m_l)/3$, that is, a weighed average of the entries

of the inverse, effective-mass tensor.[18] Table 17.4 shows that GaAs has the largest value of $1/m_n$; thus, it is expected to have the largest mobility, which is indeed the case. As far as holes are concerned, the effective masses of heavy holes of Si, Ge, and GaAs are similar to each other; also the effective masses of light holes have the same order of magnitude. Besides, considering that the valence band has two branches of $E(\mathbf{k})$, the effective masses do not combine in the simple way as for the conduction band.

The secondary minima of GaAs, placed at an energy $E_C + \Delta E$ with $\Delta E \simeq 0.36\,\text{eV}$ (Fig. 17.16), have a larger effective mass than the absolute minimum; due to this, the mobility of the electrons in the upper valleys is smaller than that of the electrons populating the absolute minimum. As ΔE is relatively small, the population of the secondary minima is not negligible; in a non-equilibrium condition, the scattering events tend increase the electron population of the upper valleys at the expense of that of the absolute minimum, with a ratio between the upper and lower population that depends on the applied electric field. This gives rise to a negative differential resistivity in the current-to-voltage curve of the material, i.e., an operating region exists where the current density decreases as the electric field increases. The phenomenon is called *Ridley–Watkins–Hilsum mechanism* [103, Sect. 14-3].

In semiconductors, the absorption of energy from an electromagnetic field may induce the transition of an electron from a state belonging to the valence band to a state belonging to the conduction band. Such a transition increases by one the number of electrons in the conduction band and, at the same time, increases by one the number of holes in the valence band; for this reason it is called *generation of an electron-hole pair*. The opposite phenomenon may also occur, namely, a release of electromagnetic energy due the transition of an electron from a state belonging to the conduction band to a state belonging to the valence band. Such a transition decreases by one the number of electrons in the conduction band and, at the same time, decreases by one the number of holes in the valence band (*recombination of an electron-hole pair*). It is worth pointing out that generation and recombination events may also occur with an energy absorption from, or release to, an external agent different from the electromagnetic field (e.g., the agent could be a vibrational mode of the lattice); for this reason, the phenomena considered here are better specified as *generations-recombinations of the radiative type*. In GaAs, the minimum of the conduction band and the maxima of the two branches of the valence band correspond to the same value of $\mathbf{k}$; semiconductors fulfilling this condition are called *direct-gap semiconductors*. Instead, Si and Ge are *indirect-gap semiconductors*, because the maxima of the valence band correspond to $\mathbf{k} = 0$, whereas the minima of the conduction band correspond to $\mathbf{k} \neq 0$. Direct- and indirect-gap semiconductors behave differently as far as generations-recombinations of the radiative type are concerned; in fact, the probability of such events is much higher in direct-gap semiconductors. This explains

[18] If the magnitudes of m_t and m_l are significantly different, the smaller effective mass dictates the magnitude of m_n.

why some classes of solid-state optical devices like, e.g., lasers, are manufactured using direct-gap semiconductors.[19]

17.6.7 Subbands

The calculations of the density of states carried out so far have been based on the assumption that all components of the **k** vector can be treated as continuous variables. In particular, the adoption of the parabolic-band approximation in the case of a periodic lattice (Sect. 17.6) leads to expressions for the density of states $g(E)$ and combined density of states $\gamma(E)$ that are formally identical to those obtained for a particle in a three-dimensional box (Sect. 15.9.2). However, in some situations it happens that not all components of **k** may be treated as continuous. To describe this case it is convenient to use the example of the box first; that of the periodic lattice is worked out later, in the frame of the parabolic-band approximation.

To proceed, consider like in Sect. 15.9.2 a three-dimensional box whose sides have lengths d_1, d_2, d_3, so that the eigenvalues of the Schrödinger equation are $E_{n_1 n_2 n_3} = \hbar^2 k^2/(2\,m)$, where k^2 is the square of

$$\mathbf{k} = n_1 \frac{\pi}{d_1}\,\mathbf{i}_1 + n_2 \frac{\pi}{d_2}\,\mathbf{i}_2 + n_3 \frac{\pi}{d_3}\,\mathbf{i}_3, \qquad n_i = 1, 2, \ldots \tag{17.85}$$

The distance between two consecutive projections of **k** along the ith side is $\Delta k_i = \pi/d_i$, and the volume associated to each **k** is $\Delta k_1\,\Delta k_2\,\Delta k_3 = \pi^3/V$, with $V = d_1\,d_2\,d_3$ the volume of the box in the **r** space. The density of the **k** vectors in the **k** space is $Q_k = V/\pi^3$.

17.6.7.1 Two-Dimensional Layer

Now, in contrast to what was implicitly assumed in Sect. 15.9.2, let one side of the box be much different from the others, for instance, $d_2 \sim d_1$, $d_3 \ll d_1, d_2$. It follows that $\Delta k_3 \gg \Delta k_1, \Delta k_2$. If the magnitudes involved are such that k_1, k_2 may still be considered continuous variables, while k_3 can not, one must calculate the density of states by treating k_1, k_2 differently from k_3. Considering $k_1 = n_1\,\pi/d_1$, $k_2 = n_2\,\pi/d_2$ as continuous, fix E and s in the relations

$$\frac{2m}{\hbar^2}\,E = k_1^2 + k_2^2 + n_3^2\,\frac{\pi^2}{d_3^2}, \qquad n_3 = s < \frac{d_3}{\pi}\,\frac{\sqrt{2m\,E}}{\hbar}. \tag{17.86}$$

[19] The reasoning seems to contradict the fact the large-area, solid-state optical sensors used in cameras and video cameras, based on the CCD or CMOS architecture, are made of silicon. In fact, the complex structure of these several-megapixel sensors and related signal-management circuitry can be realized only with the much more advanced technology of silicon. The relative ease of fabricating complex structures largely compensates for the poorer optical properties of the material.

For each integer $s = 1, 2, \ldots$ the two relations (17.86) determine in the k_1, k_2 plane a circumference of radius $c_s = \sqrt{c_s^2}$, with

$$c_s^2 = k_1^2 + k_2^2, \qquad c_s^2 = \frac{2mE}{\hbar^2} - s^2 \frac{\pi^2}{d_3^2}. \tag{17.87}$$

It is $\min_s c_s > 0$ because $s_{\max} < d_3 \sqrt{2mE}/(\pi\,\hbar)$, and $\max_s c_s = c_1 > 0$. For a fixed s the states are distributed over the circumference of radius c_s: such a set of states is also called *subband*.

The density of states in energy of the subband thus defined is calculated following the same reasoning as in Sect. 15.9.3: in fact, one observes that the density of **k** vectors in the two-dimensional space k_1, k_2 is $d_1 d_2/\pi^2$, namely, the inverse of the area $\pi^2/(d_1 d_2)$ associated to each **k** belonging to the given circumference. Then, the total number of **k** vectors in a circle of radius c_s is

$$N_{ks} = \frac{d_1\, d_2}{\pi^2}\, \pi\, c_s^2 = \frac{d_1\, d_2}{\pi} \left(\frac{2mE}{\hbar^2} - s^2 \frac{\pi^2}{d_3^2} \right). \tag{17.88}$$

Remembering that indices n_1, n_2 are positive, it is necessary to consider only the first quadrant; as a consequence, N_{ks} must be divided by 4. Further, it is necessary to multiply it by 2 to account for electron spin. In conclusion, the density of states of the two-dimensional subbands is

$$g_{2\mathrm{D}}(E) = \frac{\mathrm{d}(2\,N_{ks}/4)}{\mathrm{d}E} = \frac{d_1\, d_2 m}{\pi\,\hbar^2} = \text{const}, \tag{17.89}$$

to be compared with (15.67). Note that (17.89) is independent of index s. This result is useful, e.g., for treating the problem of a two-dimensional charge layer in the channel of a semiconductor device.

17.6.7.2 Wire

Now, assume that $d_2 \sim d_3$, and $d_2, d_3 \ll d_1$. It follows that $\Delta k_2, \Delta k_3 \gg \Delta k_1$. If the magnitudes involved are such that k_1 may still be considered a continuous variable, while k_2, k_3 can not, one must calculate the density of states by treating k_1 differently from k_2, k_3. Considering $k_1 = n_1 \pi/d_1$ as continuous, fix E, r, s in the relations

$$\frac{2m}{\hbar^2}\, E = k_1^2 + n_2^2 \frac{\pi^2}{d_2^2} + n_3^2 \frac{\pi^2}{d_3^2}, \qquad n_2 = r, \quad n_3 = s, \tag{17.90}$$

with $r^2/d_2^2 + s^2/d_3^2 \le 2mE/(\pi^2\,\hbar^2)$. For each pair of integers $r, s = 1, 2, \ldots,$ (17.90) determine in the **k** space two points given by the relation

$$\kappa_{rs}^2 = \frac{2mE}{\hbar^2} - \frac{r^2\,\pi^2}{d_2^2} - \frac{s^2\,\pi^2}{d_3^2}, \qquad \kappa_{rs} = \sqrt{\kappa_{rs}^2}. \tag{17.91}$$

It is $\min_{rs}\kappa_{rs} > 0$ and $\max_{rs}\kappa_{rs} = \kappa_{11}$. For a fixed pair r, s the states are placed at the ends of the segment $[-\kappa_{rs}, +\kappa_{rs}]$ parallel to k_1. The density of states in energy of such a segment is calculated following the same reasoning as in Sect. 15.9.3: in fact, one observes that the density of **k** vectors in the one-dimensional space k_1 is d_1/π, namely, the inverse of the length π/d_1 associated to each **k** belonging to the segment. Then, the total number of **k** vectors in the segment of length $2\,\kappa_{rs}$ is

$$N_{krs} = \frac{d_1}{\pi}\,2\kappa_{rs} = \frac{2d_1}{\pi}\left(\frac{2mE}{\hbar^2} - r^2\frac{\pi^2}{d_2^2} - s^2\frac{\pi^2}{d_3^2}\right)^{1/2}. \tag{17.92}$$

Remembering that index n_1 is non negative, it is necessary to consider only the positive half of the segment. As a consequence, N_{krs} must be divided by 2. Further, it is necessary to multiply it by 2 to account for electron spin. The density of states of the one-dimensional case then reads

$$g_{1\mathrm{D}}(E) = \frac{\mathrm{d}(2\,N_{krs}/2)}{\mathrm{d}E} = \frac{2d_1 m}{\pi\,\hbar^2\,\kappa_{rs}}, \tag{17.93}$$

to be compared with (15.69). Note that, in contrast with the two-dimensional case (17.89), here the result depends on both indices r, s. A device with $d_2, d_3 \ll d_1$ is also called *wire*. When the device size is such that the transport of a particle in it must be studied by means of Quantum Mechanics, it is also called *quantum wire*. The $E(\kappa_{rs})$ relation may be recast as

$$\frac{\hbar^2}{2\,m}\,\kappa_{rs}^2 = E - E_{rs}, \qquad E_{rs} = \frac{\pi^2\,\hbar^2}{2m}\left(\frac{r^2}{d_2^2} + \frac{s^2}{d_3^2}\right). \tag{17.94}$$

As E_{rs} is an increasing function of the indices, its minimum is attained for $r = s = 1$ and represents the ground state in the variables k_2, k_3. It is interesting to note that, if the total energy E is prescribed, e.g., by injecting the particle from an external source, such that $E_{11} < E < \min(E_{12}, E_{21})$, then the particle's wave function has the form

$$\psi = \sqrt{\frac{8}{V}}\,\sin(\kappa_{11}\,x_1)\,\sin(\pi\,x_2/d_2)\,\sin(\pi\,x_3/d_3)\,\exp(-\mathrm{i}\,E\,t/\hbar) \tag{17.95}$$

(compare with (15.60)). Remembering the expression (15.64) of the density of states in a box where $d_1 \sim d_2 \sim d_3$, the results obtained so far are summarized as:

$$g_{3\mathrm{D}}(E) = \frac{V\sqrt{2\,m^3\,E}}{\pi^2\,\hbar^3}, \quad g_{2\mathrm{D}}(E) = \frac{d_1\,d_2\,m}{\pi\,\hbar^2}, \quad g_{1\mathrm{D}}(E) = \frac{d_1\sqrt{2m}}{\pi\,\hbar\sqrt{E - E_{rs}}}. \tag{17.96}$$

17.6.8 Subbands in a Periodic Lattice

The calculations leading to (17.96) consider the case of a box within which the potential energy is zero (as a consequence, the total energy E is purely kinetic),

and prescribe a vanishing wave function at the boundaries. If a periodic lattice is present, with the provisions indicated in Sect. 17.5.3 one can apply the periodic boundary conditions. In this case, the spacing between the components of $\mathbf{k}$ in each direction doubles ($n_i\,\pi/d_i \;\leftarrow\; 2\,n_i\,\pi/d_i$), but the number of components doubles as well ($n_i = 1, 2, \ldots \leftarrow n_i = 0, \pm 1, \pm 2, \ldots$), so the density of states remains the same. In a semiconductor the calculation leading to the density of states is made more complicated by the presence of the lattice. However, the analysis may be brought to a simple generalization of that carried out in a box by means of the following simplifications:

- It is assumed that a band structure exists even if the size of the device is small in one or two spatial directions. In fact, it can be shown that the presence of a number of atomic planes of the order of ten is sufficient to form a band structure.
- The analysis is limited to the case of parabolic bands.

The case of the conduction band of silicon is considered by way of example, with the k_1, k_2, k_3 axes placed along the [100], [010], [001] directions. The parabolic-band approximation yields for the kinetic energies E_{e1}, E_{e2}, $E_{e3} \geq 0$ the expressions given in (17.82, 17.83, 17.84); the other three kinetic energies E_{e4}, E_{e5}, E_{e6} are derived from E_{e1}, E_{e2}, E_{e3}, respectively, by letting $k_m \;\leftarrow\; -k_m$. Apart from the constant E_C, the energies $E_{e1}, E_{e2}, \ldots$ are simplified forms of the eigenvalues of the Schrödinger equation (17.40). Conversely, to the purpose of determining the corresponding eigenfunctions, one may view $E_{e1}, E_{e2}, \ldots$ as the exact eigenvalues of simplified forms of the original Hamiltonian operator (17.40), that hold near the band's minima; such simplified forms are expected to be of the purely kinetic type. They are found by replacing k_i with $-\mathrm{i}\,\mathrm{d}/\mathrm{d}x_i$ in (17.82, 17.83, 17.84), this yielding

$$[100]: \quad \mathcal{H}_{e1} = \frac{\hbar^2}{2}\left[+\frac{1}{m_l}\left(\mathrm{i}\,\frac{\partial}{\partial x_1} + k_m\right)^2 - \frac{1}{m_t}\,\frac{\partial^2}{\partial x_2^2} - \frac{1}{m_t}\,\frac{\partial^2}{\partial x_3^2}\right], \tag{17.97}$$

$$[010]: \quad \mathcal{H}_{e2} = \frac{\hbar^2}{2}\left[-\frac{1}{m_t}\,\frac{\partial^2}{\partial x_1^2} + \frac{1}{m_l}\left(\mathrm{i}\,\frac{\partial}{\partial x_2} + k_m\right)^2 - \frac{1}{m_t}\,\frac{\partial^2}{\partial x_3^2}\right], \tag{17.98}$$

$$[001]: \quad \mathcal{H}_{e3} = \frac{\hbar^2}{2}\left[-\frac{1}{m_t}\,\frac{\partial^2}{\partial x_1^2} - \frac{1}{m_t}\,\frac{\partial^2}{\partial x_2^2} + \frac{1}{m_l}\left(\mathrm{i}\,\frac{\partial}{\partial x_3} + k_m\right)^2\right], \tag{17.99}$$

with m_l and m_t the longitudinal and transverse masses. Considering $\mathcal{H}_{e1}$ first, the solution of the time-independent Schrödinger equation generated by it,

$$\mathcal{H}_{e1}\,w_1(\mathbf{r}, n_1, n_2, n_3) = E(n_1, n_2, n_3)\,w_1(\mathbf{r}, n_1, n_2, n_3), \tag{17.100}$$

is found by separation, specifically, by letting $E = E_\alpha(n_1) + E_\beta(n_2) + E_\gamma(n_3)$, $w_1 = \exp(\mathrm{i}\,k_m\,x_1)\,\alpha(x_1, n_1)\,\beta(x_2, n_2)\,\gamma(x_3, n_3)$. One finds for the $\mathbf{k}$ vector the expression

$$\mathbf{k} = n_1\,\frac{\pi}{d_1}\,\mathbf{i}_1 + n_2\,\frac{\pi}{d_2}\,\mathbf{i}_2 + n_3\,\frac{\pi}{d_3}\,\mathbf{i}_3, \tag{17.101}$$

with $n_i = 1, 2, \ldots$, while the eigenfunctions and eigenvalues read

$$w_1 = \sqrt{\frac{8}{V}} \exp(\mathrm{i}\, k_m\, x_1) \sin\left(\frac{n_1\,\pi}{d_1}\, x_1\right) \sin\left(\frac{n_2\,\pi}{d_2}\, x_2\right) \sin\left(\frac{n_3\,\pi}{d_3}\, x_3\right), \quad (17.102)$$

$$E = \frac{\hbar^2}{2\,m_l}\, n_1^2\, \frac{\pi^2}{d_1^2} + \frac{\hbar^2}{2\,m_t}\, n_2^2\, \frac{\pi^2}{d_2^2} + \frac{\hbar^2}{2\,m_t}\, n_3^2\, \frac{\pi^2}{d_3^2}. \quad (17.103)$$

The eigenvalues and eigenfunctions of $\mathcal{H}_{e2}$, $\mathcal{H}_{e3}$ are found by a cyclic permutation of the indices, $1 \leftarrow 2 \leftarrow 3 \leftarrow 1$, while those of $\mathcal{H}_{e4}$, $\mathcal{H}_{e5}$, $\mathcal{H}_{e6}$ are derived from those of $\mathcal{H}_{e1}$, $\mathcal{H}_{e2}$, $\mathcal{H}_{e3}$, respectively, by letting $k_m \leftarrow -k_m$.

One notes that the $\mathbf{k}$ vectors and the eigenfunctions are not influenced by the effective masses, whereas the eigenvalues are. As a consequence, the density of states is affected as well. In the case where $d_1 \sim d_2 \sim d_3$ the density of states associated to the minimum of index 1 is found by the same procedure as that leading to the first relation in (17.96); the result is

$$g_{3\mathrm{D}}^{(1)}(E) = \frac{d_1\, d_2\, d_3\, \sqrt{2\, m_l\, m_t^2}}{\pi^2\, \hbar^3}\, \sqrt{E}. \quad (17.104)$$

Such a density of states is not affected by interchanging the effective masses; thus, the total density of states is found by adding over the densities of states of the M_C minima of the conduction band:

$$g_{3\mathrm{D}}(E) = M_C\, \frac{d_1\, d_2\, d_3\, \sqrt{2\, m_l\, m_t^2}}{\pi^2\, \hbar^3}\, \sqrt{E}. \quad (17.105)$$

As in the case of a box, the distance between two consecutive projections of $\mathbf{k}$ along the ith side is $\Delta k_i = \pi/d_i$, and the volume associated to each $\mathbf{k}$ is $\Delta k_1\, \Delta k_2\, \Delta k_3 = \pi^3/V$, with $V = d_1 d_2 d_3$. The density of the $\mathbf{k}$ vectors in the $\mathbf{k}$ space is $Q_k = V/\pi^3$.

Consider now the case of a two-dimensional layer, namely, $d_2 \sim d_1$, while $d_3 \ll d_1, d_2$. Let $k_1 = n_1\pi/d_1$, $k_2 = n_2\pi/d_2$, and fix $n_3 = 1$ whence, for the minima of indices 1 and 4,

$$E = \frac{\hbar^2}{2\, m_l}\, k_1^2 + \frac{\hbar^2}{2\, m_t}\, k_2^2 + \frac{\hbar^2}{2\, m_t}\, \frac{\pi^2}{d_3^2}, \qquad E \geq \frac{\hbar^2}{2\, m_t}\, \frac{\pi^2}{d_3^2}. \quad (17.106)$$

A calculation similar to that carried out in a box provides, for the minima of indices 1 and 4, an expression similar to that of the second relation in (17.96):

$$g_{2\mathrm{D}}^{(1)} = g_{2\mathrm{D}}^{(4)} = \frac{d_1\, d_2 \sqrt{m_l\, m_t}}{\pi\, \hbar^2}. \quad (17.107)$$

For the other pairs of minima one finds

$$g_{2\mathrm{D}}^{(2)} = g_{2\mathrm{D}}^{(5)} = \frac{d_1\, d_2\, \sqrt{m_l\, m_t}}{\pi\, \hbar^2}, \qquad g_{2\mathrm{D}}^{(3)} = g_{2\mathrm{D}}^{(6)} = \frac{d_1\, d_2\, m_t}{\pi\, \hbar^2}. \quad (17.108)$$

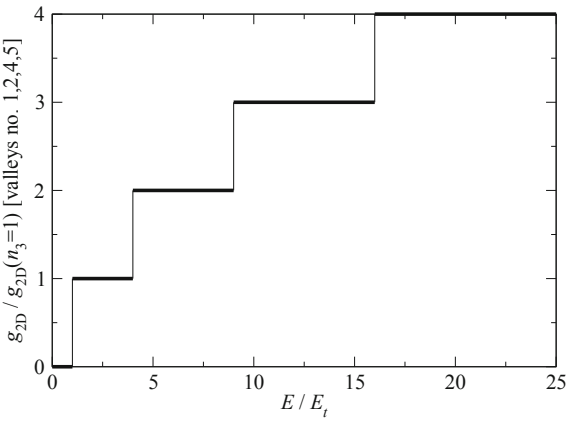

Fig. 17.19 Normalized, two-dimensional density of states (17.109) for the 1, 2, 4, 5 valleys of silicon, as a function of E/E_t, in the parabolic-band approximation

In conclusion, for a two-dimensional layer with $d_3 \ll d_1, d_2$ and $n_3 = 1$, within the parabolic-band approximation, the density of states for the minima of indices 1, 2, 4, and 5 is the same constant for all energies $E \geq \hbar^2 \pi^2/(2\, m_t\, d_3^2)$. The total density of states for these minima is

$$g_{2D}^{(1,2,4,5)} = 4\, \frac{d_1\, d_2\, \sqrt{m_l\, m_t}}{\pi\, \hbar^2}. \qquad E \geq E_t = \frac{\hbar^2\, \pi^2}{2\, m_t\, d_3^2}, \qquad n_3 = 1, \tag{17.109}$$

while $g_{2D,n_3=1}^{(1,2,4,5)} = 0$ for $E < E_t$. Similarly, still with $n_3 = 1$, the density of states for the minima of indices 3 and 6 is another constant for all energies $E \geq \hbar^2\, \pi^2/(2\, m_l\, d_3^2)$. The total density of states for these minima is

$$g_{2D}^{(3,6)} = 2\, \frac{d_1\, d_2\, m_t}{\pi\, \hbar^2}, \qquad E \geq E_l = \frac{\hbar^2\, \pi^2}{2\, m_l\, d_3^2}, \qquad n_3 = 1, \tag{17.110}$$

while $g_{2D,n_3=1}^{(3,6)} = 0$ for $E < E_l$. Now, let $n_3 = 2$; it is easily found that the value of $g_{2D,n_3=2}^{(1,2,4,5)}$ is the same as above, however, it holds for $E \geq 4\, E_t$. It adds up to the value found for $n_3 = 1$, giving rise to a stair-like form of $g_{2D}^{(1,2,4,5)}$ as a function of energy. The same is obtained for $g_{2D,n_3=2}^{(3,6)}$ when $E \geq 4\, E_l$, and so on. An example of such a density of states is sketched in Fig. 17.19, where the ratio $g_{2D}^{(1,2,4,5)}/g_{2D,n_3=1}^{(1,2,4,5)}$ is shown as a function of E/E_t. The total density of states is found by adding up the two stair-like functions. From Table 17.4 one finds that in silicon at room temperature it is $m_l \simeq 0.97\, m_0$, $m_t \simeq 0.19\, m_0$, whence $E_t \simeq 5.1\, E_l$ and $g_{2D,n_3=1}^{(1,2,4,5)} \simeq 4.47\, g_{2D,n_3=1}^{(3,6)}$.

As shown by Fig. 17.19, the derivative of the density of states with respect to energy diverges at some points. Such divergences are called *Van Hove singularities* [2, Chap. 8].

Finally, consider the case of a wire, namely, $d_2 \sim d_3$, while $d_2, d_3 \ll d_1$. Let $k_1 = n_1\pi/d_1$ and fix $n_2 = n_3 = 1$ whence, for the minima of indices 1 and 4,

$$E = \frac{\hbar^2}{2\,m_l}\,k_1^2 + \frac{\hbar^2}{2\,m_t}\,\frac{\pi^2}{d_2^2} + \frac{\hbar^2}{2\,m_t}\,\frac{\pi^2}{d_3^2}, \qquad E \geq \frac{\pi^2\,\hbar^2}{2\,m_t}\left(\frac{1}{d_2^2} + \frac{1}{d_3^2}\right) = E_{11}^{(1,4)}. \tag{17.111}$$

A calculation similar to that carried out in a box provides, for the minima of indices 1 and 4, an expression similar to that of the third relation in (17.96):

$$g_{1D}^{(1)} = g_{1D}^{(4)} = \frac{2\,d_1\,m_l}{\pi\,\hbar^2\,\kappa_{11}^{(1,4)}} = \frac{d_1\,\sqrt{2\,m_l}}{\pi\,\hbar\,\sqrt{E - E_{11}^{(1,4)}}}, \tag{17.112}$$

while $g_{1D}^{(1,4)} = 0$ if $E < E_{11}^{(1,4)}$. For the other minima one finds

$$g_{1D}^{(2,5)} = \frac{d_1\,\sqrt{2\,m_t}}{\pi\,\hbar\,\sqrt{E - E_{11}^{(2,5)}}}, \qquad E \geq E_{11}^{(2,5)} = \frac{\pi^2\,\hbar^2}{2}\left(\frac{1}{m_l\,d_2^2} + \frac{1}{m_t\,d_3^2}\right), \tag{17.113}$$

with $g_{1D}^{(2,5)} = 0$ if $E < E_{11}^{(2,5)}$, and

$$g_{1D}^{(3,6)} = \frac{d_1\,\sqrt{2\,m_t}}{\pi\,\hbar\,\sqrt{E - E_{11}^{(3,6)}}}, \qquad E \geq E_{11}^{(3,6)} = \frac{\pi^2\,\hbar^2}{2}\left(\frac{1}{m_t\,d_2^2} + \frac{1}{m_l\,d_3^2}\right), \tag{17.114}$$

with $g_{1D}^{(3,6)} = 0$ if $E < E_{11}^{(3,6)}$. In conclusion, for a wire with $d_2, d_3 \ll d_1$ and $n_2 = n_3 = 1$, within the parabolic-band approximation, the density of states of each pair of minima is the sum of expressions of the form (17.112, 17.113, 17.114); the latter are complicate because all possible pairs of indices r, s combine with the two lengths d_2, d_3, that in general are not commensurable with each other. A somewhat easier description is obtained by considering (17.112) alone and letting $d_2 = d_3$ in it; this yields $E_{21} = E_{12} = 2.5\,E_{11}$, $E_{22} = 4\,E_{11}$, $E_{31} = E_{13} = 5\,E_{11}$, and so on, and

$$g_{1D}^{(1,4)} = \frac{g_{1D}^{(1,4)}(E = 2\,E_{11})}{\sqrt{E/E_{11} - 1}}, \qquad E_{11} < E \leq E_{21}. \tag{17.115}$$

In the next interval $E_{21} < E \leq E_{22}$, the density of states is the sum of (17.115) and $2/\sqrt{E/E_{11} - 2.5}$, where factor 2 accounts for the $(r = 2, s = 1)$, $(r = 1, s = 2)$ degeneracy; in the interval $E_{22} < E \leq E_{31}$ one adds the further summand $1/\sqrt{E/E_{11} - 4}$, and so on. The normalized density of states $g_{1D}(E)/g_{1D}(2\,E_{11})$ is shown in Fig. 17.20 as a function of E/E_{11}.

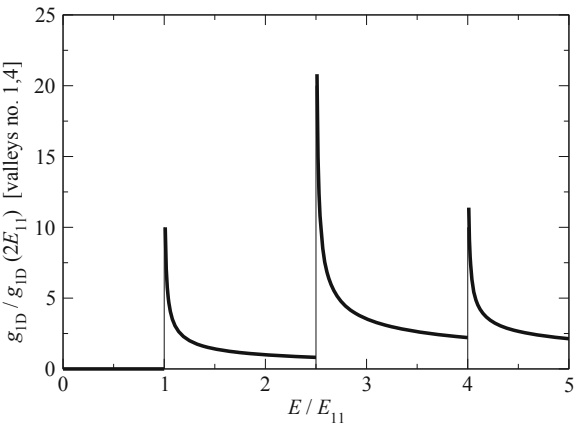

Fig. 17.20 Normalized, one-dimensional density of states for the 1, 4 valleys of silicon, as a function of E/E_{11}, in the parabolic-band approximation and with $d_2 = d_3$

Also in this case the Van Hove singularities are present; in addition, the density of states itself diverges at such points. However, such divergences are integrable; consider for instance an integral of the form

$$\int_{E_0}^{\infty} \frac{c}{\sqrt{E - E_0}} P(E)\,\mathrm{d}E, \tag{17.116}$$

with c a constant and $0 < P < 1$ a distribution function. Splitting the integration domain into two intervals $E_0 \le E \le E'$ and $E' \le E < \infty$, with $E' > E_0$, one finds for the first integral, that contains the singularity,

$$\int_{E_0}^{E'} \frac{c}{\sqrt{E - E_0}} P(E)\,\mathrm{d}E \le \int_{E_0}^{E'} \frac{c}{\sqrt{E - E_0}}\,\mathrm{d}E < \infty. \tag{17.117}$$

17.7 Calculation of Vibrational Spectra

The discussion carried out in Sect. 16.6 has led to the conclusion that in the case of solid matter the nuclei, being massive and tightly bound together, are expected to depart little from their equilibrium positions $\mathbf{R}_0$. The classical description of the nuclear motion is thus brought to the case already solved in Sects. 3.9 and 3.10: the vibrational state of the nuclei is described in terms of the normal coordinates b_σ, whose conjugate moments are $\dot{b}_\sigma$, and the total energy of the nuclei reads (compare with (3.50))

$$T_a + V_a = \sum_{\sigma=1}^{3N} H_\sigma + V_{a0}, \qquad H_\sigma = \frac{1}{2}\,\dot{b}_\sigma^2 + \frac{1}{2}\,\omega_\sigma^2\, b_\sigma^2, \tag{17.118}$$

where each H_σ corresponds to one degree of freedom and $\omega_\sigma > 0$ is the angular frequency of the corresponding mode. The system is completely separable in the

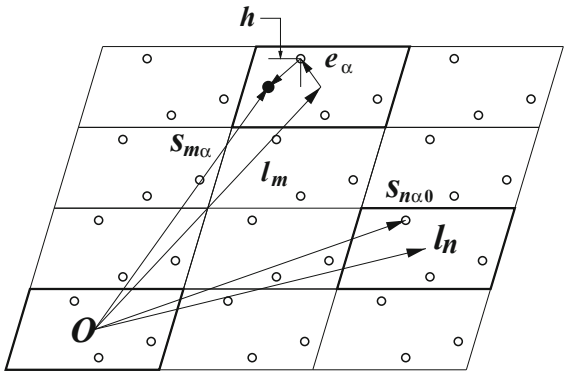

Fig. 17.21 Definition of the labels used to identify the degrees of freedom in a periodic lattice

normal coordinates, and each normal coordinate evolves in time as a linear harmonic oscillator. The calculation is based on Classical Mechanics; it is carried out in this chapter because it exploits the periodicity properties of the material and, in this respect, presents several analogies with the solution of the Schrödinger equation in a periodic lattice. To determine the vibrational frequencies ω_σ it is necessary to solve the eigenvalue equation (3.43), namely,

$$\mathbf{C}\,\mathbf{g}_\sigma = \omega_\sigma^2\,\mathbf{M}\,\mathbf{g}_\sigma, \qquad \sigma = 1, \ldots, 3\,N, \tag{17.119}$$

with $\mathbf{g}_\sigma$ the eigenvectors. The entries of $\mathbf{C}$, $\mathbf{M}$ are given by

$$c_{kn} = [\mathbf{C}]_{kn} = \left(\frac{\partial^2 V_a}{\partial h_k\,\partial h_n}\right)_0, \qquad [\mathbf{M}]_{kn} = \mu_n\,\delta_{kn}, \tag{17.120}$$

where V_a is the potential energy, h_k the displacement of the kth degree of freedom with respect to the equilibrium position, μ_k the mass associated to the kth degree of freedom, and δ_{kn} the Kronecker delta.

The calculation is in principle the same for any system of particles; however, if the system has special properties, they reflect into the form of the eigenvalues and eigenvectors. A particularly important case is that of a periodic structure, such as a crystal. Considering this case, let the crystal be made of N_c elementary cells, with a basis made of N_b nuclei (the definition of basis is in Sect. 17.2). It follows that the total number of nuclei is $N = N_b\,N_c$, and the total number of degrees of freedom is $3\,N$. With respect to a given origin O (Fig. 17.21), the mth cell of the lattice is identified by the corresponding translation vector of the direct lattice, $\mathbf{l}_m$; the latter determines a local origin within the mth cell. In turn, the equilibrium position of the αth nucleus of the mth cell with respect to the local origin is identified by a vector $\mathbf{e}_\alpha$ of the direct lattice.

17.7.1 Labeling the Degrees of Freedom— Dynamic Matrix

To proceed, it is convenient to label the degrees of freedom in such a way as to distinguish the indices of the cells from those of the basis and of the coordinate axes. To this purpose, one observes that the component along the uth coordinate axis of the equilibrium position of the jth nucleus is

$$X_{ju0} = s_{q0}, \qquad q = u + 3\,(j-1), \qquad j = \alpha + N_b\,(m-1), \tag{17.121}$$

with $u = 1,2,3$; $\alpha = 1,\ldots,N_b$; $m = 1,\ldots,N_c$. The same applies to the displacements, which are more conveniently expressed in terms of three indices:

$$h_q \longleftarrow h_{m\alpha u}, \qquad h_r \longleftarrow h_{n\beta w}. \tag{17.122}$$

The entries of $\mathbf{C}$ are identified in the same manner:

$$c_{qr} = \left(\frac{\partial^2 V_a}{\partial h_q\,\partial h_r}\right)_0 \qquad \longleftarrow \qquad c_{m\alpha u}^{n\beta w} = \left(\frac{\partial^2 V_a}{\partial h_{m\alpha u}\,\partial h_{n\beta w}}\right)_0, \tag{17.123}$$

with

$$m,n = 1,\ldots,N_c, \qquad \alpha,\beta = 1,\ldots,N_b, \qquad u,w = 1,2,3. \tag{17.124}$$

The order of derivation is irrelevant, so that $c_{m\alpha u}^{n\beta w} = c_{n\beta w}^{m\alpha u}$. As the number of nuclei is finite, the crystal is not actually periodic; as indicated above, periodicity is recovered by imposing periodic boundary conditions to the quantities of interest (Sect. 17.5.3).[20] With this provision, the entries of $\mathbf{C}$ are invariant with respect to the lattice translations. The latter are related only to the cell indices m,n and are obtained by the replacements $\mathbf{l}_m \leftarrow \mathbf{l}_m + \mathbf{l}_\nu$, $\mathbf{l}_n \leftarrow \mathbf{l}_n + \mathbf{l}_\nu$, with ν any integer. In particular, taking $\mathbf{l}_\nu = -\mathbf{l}_n$ yields

$$c_{m\alpha u}^{n\beta w} = c_{\alpha u}^{\beta w}(\mathbf{l}_m, \mathbf{l}_n) = c_{\alpha u}^{\beta w}(\mathbf{l}_m - \mathbf{l}_n, 0) = c_{\alpha u}^{\beta w}(\mathbf{l}_m - \mathbf{l}_n). \tag{17.125}$$

The above shows that the entries of $\mathbf{C}$ depend on the relative positions of the cells. Due to the invariance of $\mathbf{C}$ with respect to the lattice translations one sees that, given α, u and β, w, there are only N_c distinct entries of $\mathbf{C}$ out of N_c^2, namely, the distinct entries are those such that $m-n = 0, m-n = 1, \ldots, m-n = N_c - 1$. In fact, all remaining $N_c^2 - N_c$ entries are derived from the first N_c ones by suitable translations of the indices. In turn, using the new indices (17.124) the entries of $\mathbf{M}$ read

$$\mu_r\,\delta_{qr} \qquad \longleftarrow \qquad \mu_{n\beta w}\,\delta_{m\alpha u}^{n\beta w} = \mu_\beta\,\delta_{m\alpha u}^{n\beta w}, \tag{17.126}$$

[20] As mentioned in Sect. 17.5.3, the periodic boundary conditions are actually an approximation; however, the interatomic interactions typically give rise to short-range forces, hence the above reasoning holds for all the cells that are not too close to the boundaries.

where the last equality is due to the fact that the mass of a given nucleus of the cell does not depend on the cell position within the crystal nor on the coordinate axis. In the new indices the eigenvalue equation (17.119) becomes

$$\sum_{n\beta w} c^{n\beta w}_{m\alpha u}\, g_{n\beta w} = \omega^2\, \mu_\alpha\, g_{m\alpha u}, \tag{17.127}$$

where the indices' ranges are given in (17.124). The indices of the eigenvalue and eigenvector have been omitted for simplicity. Defining

$$d^{n\beta w}_{m\alpha u} = \frac{c^{n\beta w}_{m\alpha u}}{\sqrt{\mu_\alpha\, \mu_\beta}}, \qquad z_{m\alpha u} = \sqrt{\mu_\alpha}\, g_{m\alpha u}, \qquad z_{n\beta w} = \sqrt{\mu_\beta}\, g_{n\beta w}, \tag{17.128}$$

transforms (17.127) into

$$\sum_{n\beta w} d^{n\beta w}_{m\alpha u}\, z_{n\beta w} = \omega^2\, z_{m\alpha u}. \tag{17.129}$$

The latter form of the eigenvalue equation is more convenient because it eliminates the coefficient μ_α from the right hand side. Matrix $\mathbf{D}$ of entries $d^{n\beta w}_{m\alpha u}$ is called *dynamic matrix* and, due to the properties of $\mathbf{C}$, is symmetric ($d^{n\beta w}_{m\alpha u} = d^{m\alpha u}_{n\beta w}$) and translationally invariant:

$$d^{n\beta w}_{m\alpha u} = d^{\beta w}_{\alpha u}(\mathbf{l}_m, \mathbf{l}_n) = d^{\beta w}_{\alpha u}(\mathbf{l}_m - \mathbf{l}_n, 0) = d^{\beta w}_{\alpha u}(\mathbf{l}_m - \mathbf{l}_n). \tag{17.130}$$

17.7.2 Application of the Bloch Theorem

As a consequence of the translational invariance of $\mathbf{D}$, Bloch's theorem (17.23) applies,[21] namely, for any eigenvector of indices $k\gamma e$, and letting $\mathbf{l}_0 = 0$, the following holds:

$$\mathbf{z}_{\gamma e}(\mathbf{l}_k) = \exp(\mathbf{c} \cdot \mathbf{l}_k)\, \mathbf{z}_{\gamma e}(0). \tag{17.131}$$

In (17.131), $\mathbf{c}$ is any complex vector of the reciprocal lattice, and $k = 0, \ldots, N_c - 1$; $\gamma = 1, \ldots, N_b$; $e = 1, 2, 3$. The complex form of the eigenvectors is adopted for convenience; at the end of the calculation, a set of real eigenvectors is recovered from suitable combinations of the complex ones. Using the periodic boundary conditions, the expression of $\mathbf{c}$ is found to be

$$\mathbf{c} = \mathrm{i}\, \mathbf{q}, \qquad \mathbf{q} = \sum_{s=1}^{3} \frac{\nu_s}{N_s}\, 2\pi\, \mathbf{b}_s, \tag{17.132}$$

[21] The Bloch theorem was derived in Sect. 17.5.1 with reference to the eigenfunctions of a translation operator in the continuous case; the theorem equally holds for a translation operator in the discrete case, like the dynamic matrix considered here.

with N_1, N_2, N_3 the number of cells along the directions of the characteristic vectors of the direct lattice, $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ the characteristic vectors of the reciprocal lattice, and ν_1, ν_2, ν_3 integers, with $\nu_s = 0, 1, \ldots, N_s - 1$. The total number of distinct $\mathbf{q}$ vectors is thus $N_1\, N_2\, N_3 = N_c$. Comparing (17.132) with (17.37) shows that the structure of the $\mathbf{q}$ vector is the same as that of the $\mathbf{k}$ vector found in the solution of the Schrödinger equation (Sect. 17.5.3). Inserting (17.130) into (17.129) yields, for the line of indices $m\alpha u$ of the eigenvalue equation,

$$\sum_{n\beta w} A^{n\beta w}_{m\alpha u}\, z_{\beta w}(0) = \omega^2 z_{\alpha u}(0), \tag{17.133}$$

with

$$A^{n\beta w}_{m\alpha u} = \frac{1}{\sqrt{\mu_\alpha \mu_\beta}}\, c^{\beta w}_{\alpha u}(\mathbf{l}_m - \mathbf{l}_n)\, \exp\left[\mathrm{i}\, \mathbf{q} \cdot (\mathbf{l}_n - \mathbf{l}_m)\right]. \tag{17.134}$$

As the eigenvalues ω^2 are real, the matrix made of the entries $A^{n\beta w}_{m\alpha u}$ must be Hermitean; in fact, this is easily found by observing that $\mathbf{D}$ is real and symmetric:

$$A^{m\alpha u}_{n\beta w} = d^{\alpha u}_{\beta w}(\mathbf{l}_n - \mathbf{l}_m)\, \exp\left[\mathrm{i}\, \mathbf{q} \cdot (\mathbf{l}_m - \mathbf{l}_n)\right] = \left(A^{n\beta w}_{m\alpha u}\right)^*. \tag{17.135}$$

Another property stems from the expression at the left hand side of (17.133),

$$\sum_{n\beta w} A^{n\beta w}_{m\alpha u}\, z_{\beta w}(0) = \sum_{\beta w} \left(\sum_n A^{n\beta w}_{m\alpha u} \right) z_{\beta w}(0), \tag{17.136}$$

where $A^{n\beta w}_{m\alpha u}$ is translationally invariant because it depends on the cell indices only through the difference $\mathbf{l}_m - \mathbf{l}_n$. It follows that $\sum_n A^{n\beta w}_{m\alpha u}$ does not depend on m. This is easily verified by carrying out the sum first with, say, $m = 1$, then with $m = 2$, and observing that the terms of the second sum are the same as in the first one, displaced by one position. In summary, letting $\mathbf{A}$ be the $3\, N_b \times 3\, N_b$, Hermitean matrix of entries

$$A^{\beta w}_{\alpha u}(\mathbf{q}) = \sum_{n=1}^{N_c} d^{\beta w}_{\alpha u}(\mathbf{l}_m - \mathbf{l}_n)\, \exp\left[\mathrm{i}\, \mathbf{q} \cdot (\mathbf{l}_n - \mathbf{l}_m)\right], \tag{17.137}$$

(17.133) becomes

$$\sum_{\beta w} A^{\beta w}_{\alpha u}(\mathbf{q})\, z_{\beta w}(0) = \omega^2 z_{\alpha u}(0). \tag{17.138}$$

For a given $\mathbf{q}$, (17.138) is an eigenvalue equation of order $3\, N_b$, whose eigenvalues are found by solving the algebraic equation

$$\det(\mathbf{A} - \omega^2\, \mathbf{I}) = 0, \tag{17.139}$$

with $\mathbf{I}$ the identity matrix. As the entries of $\mathbf{A}$ depend on $\mathbf{q}$, the calculation of the $3\,N_b$ eigenvalues of (17.138) must be repeated for each distinct value of $\mathbf{q}$, namely, N_c times. The total number of eigenvalues thus found is $3\,N_b \times N_c = 3\,N$, as should be. This result shows that, while the translational invariance eliminates the dependence on $\mathbf{l}_m$, it introduces that on $\mathbf{q}$. As the number of different determinations of the two vectors $\mathbf{l}_m$ and $\mathbf{q}$ is the same, namely, N_c, the total number of eigenvalues is not affected. Letting the N_c determinations of $\mathbf{q}$ be numbered as $\mathbf{q}_1, \mathbf{q}_2 \ldots \mathbf{q}_k \ldots$, the algebraic system (17.138) is recast as

$$\sum_{\beta w} A^{\beta w}_{\alpha u}(\mathbf{q}_k)\, z_{\beta w}(0, \mathbf{q}_k) = \omega^2(\mathbf{q}_k)\, z_{\alpha u}(0, \mathbf{q}_k), \qquad k = 1, 2, \ldots N_c \tag{17.140}$$

which, for each $\mathbf{q}_k$, yields $3\,N_b$ eigenvalues ω^2 and $3\,N_b$ eigenvectors of length $3\,N_b$; as a consequence, the set of column vectors made of the eigenvectors associated to $\mathbf{q}_k$ forms a $3\,N_b \times 3\,N_b$ matrix, indicated here with $\mathbf{Z}_{1k}$. By letting $\mathbf{q}_k$ span over all its N_c determinations, the total number of eigenvalues turns out to be $3\,N$, namely,

$$\omega^2_{\gamma e}(\mathbf{q}_1), \ldots, \omega^2_{\gamma e}(\mathbf{q}_{N_c}), \qquad \gamma = 1, \ldots, N_b, \qquad e = 1, 2, 3. \tag{17.141}$$

Similarly, the total number of eigenvectors (of order $3\,N_b$) turns out to be $3\,N$,

$$\mathbf{z}_{\gamma e}(0, \mathbf{q}_1), \ldots, \mathbf{z}_{\gamma e}(0, \mathbf{q}_{N_c}), \quad \gamma = 1, \ldots, N_b, \quad e = 1, 2, 3. \tag{17.142}$$

They provide the set of N_c square matrices of order $3\,N_b$, indicated with $\mathbf{Z}_{11}, \mathbf{Z}_{12}, \ldots, \mathbf{Z}_{1N_c}$. Finally, each $\mathbf{z}_{\gamma e}(0, \mathbf{q}_k)$ provides an eigenvector of order $3\,N$ whose entries are

$$z^{\alpha u}_{\gamma e}(\mathbf{l}_m, \mathbf{q}_k) = \exp(\mathrm{i}\,\mathbf{q}_k \cdot \mathbf{l}_m)\; z^{\alpha u}_{\gamma e}(0, \mathbf{q}_k), \tag{17.143}$$

where, as usual, $\alpha, \gamma = 1, \ldots, N_b$; $u, e = 1, 2, 3$ and, in turn, $m = 0, \ldots, N_c - 1$; $k = 1, \ldots, N_c$. The first index of matrices $\mathbf{Z}_{11}, \mathbf{Z}_{12}, \ldots$ corresponds to $m = 0$. Similarly, index $m = 1$ provides a new set of matrices $\mathbf{Z}_{21}, \mathbf{Z}_{22}, \ldots$, and so on. The whole set of N_c^2 matrices $\mathbf{Z}_{mk}$ is equivalent to the $3\,N \times 3\,N$ matrix $\mathbf{Z}$ of the eigenvectors of the dynamic matrix, according to the following scheme:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_{11} & \mathbf{Z}_{12} & \ldots & \mathbf{Z}_{1N_c} \\ \mathbf{Z}_{21} & \mathbf{Z}_{22} & \ldots & \mathbf{Z}_{2N_c} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_{N_c 1} & \mathbf{Z}_{N_c 2} & \ldots & \mathbf{Z}_{N_c N_c} \end{bmatrix}. \tag{17.144}$$

17.7.3 Properties of the Eigenvalues and Eigenvectors

Remembering that $\mathbf{A}$ (defined in (17.137)) is Hermitean, and ω^2 is real, one finds $(\mathbf{A} - \omega^2\,\mathbf{I})^* = \mathbf{A}^* - \omega^2\,\mathbf{I} = \mathbf{A}^T - \omega^2\,\mathbf{I} = (\mathbf{A} - \omega^2\,\mathbf{I})^T$, whence

$$\det[(\mathbf{A} - \omega^2\,\mathbf{I})^*] = \det[(\mathbf{A} - \omega^2\,\mathbf{I})^T] = \det(\mathbf{A} - \omega^2\,\mathbf{I}). \tag{17.145}$$

This shows that the eigenvalue equation $\mathbf{A}(\mathbf{q}_k)\,\mathbf{z}(0,\mathbf{q}_k) = \omega^2(\mathbf{q}_k)\,\mathbf{z}(0,\mathbf{q}_k)$, and its conjugate, $\mathbf{A}^*(\mathbf{q}_k)\,\mathbf{z}^*(0,\mathbf{q}_k) = \omega^2(\mathbf{q}_k)\,\mathbf{z}^*(0,\mathbf{q}_k)$ have the same eigenvalues. Moreover, as the entries (17.137) of $\mathbf{A}$ are polynomials in $\exp[\mathrm{i}\,\mathbf{q}_k\cdot(\mathbf{l}_n-\mathbf{l}_m)]$ with real coefficients, the following hold:

$$A^{\beta w}_{\alpha u}(-\mathbf{q}_k) = \left[A^{\beta w}_{\alpha u}(\mathbf{q}_k)\right]^*, \qquad \mathbf{A}(-\mathbf{q}_k) = \mathbf{A}^*(\mathbf{q}_k). \tag{17.146}$$

The above properties give rise to other important consequences for the eigenvalues and eigenvectors. In fact, from the property $\mathbf{A}(-\mathbf{q}_k) = \mathbf{A}^*(\mathbf{q}_k)$ and the hermiticity of $\mathbf{A}$ one finds

$$\det[\mathbf{A}(-\mathbf{q}_k) - \omega^2\,\mathbf{I}] = \det\{[\mathbf{A}(\mathbf{q}_k) - \omega^2\,\mathbf{I}]^T\} = \det[\mathbf{A}(\mathbf{q}_k) - \omega^2\,\mathbf{I}], \tag{17.147}$$

showing that the eigenvalues calculated from $\mathbf{A}(-\mathbf{q}_k)$ are the same as those calculated from $\mathbf{A}(\mathbf{q}_k)$. It follows that ω is an even function of $\mathbf{q}_k$:

$$\omega(-\mathbf{q}_k) = \omega(\mathbf{q}_k), \qquad \mathbf{A}(-\mathbf{q}_k)\,\mathbf{z}(0,-\mathbf{q}_k) = \omega^2(\mathbf{q}_k)\,\mathbf{z}(0,-\mathbf{q}_k). \tag{17.148}$$

Taking the conjugate of the second equation in (17.148) and using again the relation $\mathbf{A}(-\mathbf{q}_k) = \mathbf{A}^*(\mathbf{q}_k)$ yields $\mathbf{A}(\mathbf{q}_k)\,\mathbf{z}^*(0,-\mathbf{q}_k) = \omega^2(\mathbf{q}_k)\,\mathbf{z}^*(0,-\mathbf{q}_k)$. Comparing the above with the original eigenvalue equation $\mathbf{A}(\mathbf{q}_k)\,\mathbf{z}(0,\mathbf{q}_k) = \omega^2(\mathbf{q}_k)\,\mathbf{z}(0,\mathbf{q}_k)$ provides a relation between the eigenvectors:

$$\mathbf{z}(0,-\mathbf{q}) = \mathbf{z}^*(0,\mathbf{q}). \tag{17.149}$$

From Bloch's theorem (17.131) it follows $z^{\alpha u}_{\gamma e}(\mathbf{l}_m,\mathbf{q}_k) = \exp(\mathrm{i}\,\mathbf{q}_k\cdot\mathbf{l}_m)\,z^{\alpha u}_{\gamma e}(0,\mathbf{q}_k)$ which, combined with (17.149), allows one to recover a set of real eigenvectors of the dynamic matrix:

$$z^{\alpha u}_{\gamma e}(\mathbf{l}_m,\mathbf{q}_k) + z^{\alpha u}_{\gamma e}(\mathbf{l}_m,-\mathbf{q}_k) = z^{\alpha u}_{\gamma e}(0,\mathbf{q}_k)\,\exp(\mathrm{i}\,\mathbf{q}_k\cdot\mathbf{l}_m) + z^{*\,\alpha u}_{\gamma e}(0,\mathbf{q}_k)\,\exp(-\mathrm{i}\,\mathbf{q}_k\cdot\mathbf{l}_m)$$

where, as usual, the indices $k\gamma e$ count the eigenvectors and the indices $m\alpha u$ count the entries. Using the results of Sect. 3.10, the displacements of the particles from the equilibrium position are given by $\mathbf{h} = \mathbf{G}\mathbf{b}$, where $\mathbf{G}$ is the matrix of the eigenvalues of (17.119) and the entries of $\mathbf{b}$ have the form (3.49), namely,

$$b_{k\gamma e}(t) = \frac{1}{2}\left\{\tilde{b}_{k\gamma e0}\,\exp[-\mathrm{i}\,\omega_{\gamma e}(\mathbf{q}_k)\,t] + \tilde{b}^*_{k\gamma e0}\,\exp[\mathrm{i}\,\omega_{\gamma e}(\mathbf{q}_k)\,t]\right\}, \tag{17.150}$$

with $\tilde{b}_{k\gamma e0}$ depending on the initial conditions $b_{k\gamma e0}(0)$, $\dot{b}_{k\gamma e0}(0)$. In turn, the entries of matrix $\mathbf{g}$ are $g^{m\alpha u}_{k\gamma e}$, where the lower indices refer to the columns and count the eigenvectors, the upper ones refer to the rows and count the entries of each eigenvector. Due to (17.128), such entries equal the corresponding terms of the real eigenvector of the dynamic matrix, divided by $\sqrt{\mu_\alpha}$. In conclusion, from $\mathbf{h} = \mathbf{G}\mathbf{b}$, the displacements are given by

$$h_{m\alpha u} = \sum_{k\gamma e} g^{m\alpha u}_{k\gamma e}\,b_{k\gamma e} = \sum_{k\gamma e}\frac{1}{\sqrt{\mu_\alpha}}\,z^{m\alpha u}_{k\gamma e}\,b_{k\gamma e}. \tag{17.151}$$

Using (17.150) yields

$$h_{m\alpha u} = \frac{1}{\sqrt{\mu_\alpha}} \Re \sum_{k\gamma e} z^{\alpha u}_{\gamma e}(0, \mathbf{q}_k) \left[\tilde{b}_{k\gamma e0} \exp(\mathrm{i}\, \Phi^m_{k\gamma e}) + \tilde{b}^*_{k\gamma e0} \exp(\mathrm{i}\, \Psi^m_{k\gamma e}) \right], \tag{17.152}$$

where the phases are defined by

$$\Phi^m_{k\gamma e} = \mathbf{q}_k \cdot \mathbf{l}_m - \omega_{\gamma e}(\mathbf{q}_k)\, t, \qquad \Psi^m_{k\gamma e} = \mathbf{q}_k \cdot \mathbf{l}_m + \omega_{\gamma e}(\mathbf{q}_k)\, t. \tag{17.153}$$

The above result shows that, in the harmonic approximation, the displacements have the form of a superposition of plane and monochromatic waves, whose wave vector and angular frequency are $\mathbf{q}_k$, $\omega_{\gamma e}(\mathbf{q}_k)$. The wave corresponding to a given $\mathbf{q}_k$ is called *vibrational mode*. Typically, the number of $\mathbf{q}_k$ vectors is very large; in such cases, the same reasoning made in Sect. 17.6 with reference to the $\mathbf{k}$ vectors holds, and $\mathbf{q}$ is considered a continuous variable ranging over the first Brillouin zone.

Function $\omega_{\gamma e}(\mathbf{q})$ is also called *dispersion relation*, and is viewed as a multi-valued function of $\mathbf{q}$ having $3N_b$ branches. For each branch, letting $q = |\mathbf{q}|$, the *wavelength* is defined by $\lambda = 2\pi/q$ and the *phase velocity* by $u_f = \omega/q = \lambda\, \nu$, with $\nu = \omega/(2\pi)$ the frequency.

The *group velocity* is defined by $\mathbf{u} = \mathrm{grad}_{\mathbf{q}}\omega$. As shown in Sect. 3.10, the total energy of the system is the sum of the mode energies, and in the classical description is expressed in terms of the initial conditions as

$$T_a + V_a = V_{a0} + \sum_{\sigma=1}^{3N} E_\sigma, \qquad E_\sigma = \frac{1}{2}\, \dot{b}^2_\sigma(0) + \frac{1}{2}\, \omega^2_\sigma\, b^2_\sigma(0). \tag{17.154}$$

As remarked in Sect. 12.5, the classical expression of the energy associated to each mode has the same form as that of a mode of the electromagnetic field. In turn, the energy quantization shows that each mode energy is made of terms of the form $\hbar\omega_{\gamma e}(\mathbf{q})$, this leading to the concept of phonon (Eqs. (12.35, 12.36)).

17.8 Complements

17.8.1 Crystal Planes and Directions in Cubic Crystals

From the general definition (17.1) of the translation vector, that provides the positions of all nodes of the crystal, it follows that a *crystal plane* is defined by the set of all triads of integers m_1, m_2, m_3 such that, for a given vector $\mathbf{g}_0$ of the scaled reciprocal lattice, the quantity $(m_1\, \mathbf{a}_1 + m_2\, \mathbf{a}_2 + m_3\, \mathbf{a}_3) \cdot \mathbf{g}_0/(2\pi)$ equals a fixed integer. Such a plane is normal to $\mathbf{g}_0$. In turn, given two crystal planes defined as above using, respectively, two non-parallel vectors $\mathbf{g}_1$ and $\mathbf{g}_2$, a *crystal direction* is defined by the set of all triads of integers m_1, m_2, m_3 that belong to the two crystal planes so

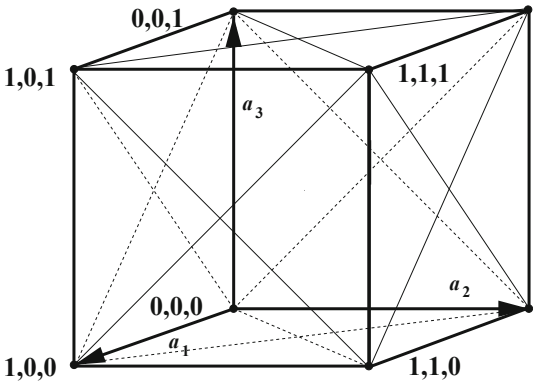

Fig. 17.22 Example of node labeling in the cubic lattice

prescribed. In cubic crystals, the typical method by which the crystal planes are identified is outlined below [103, Sect. 2.2].

Let the plane be indicated with Π. After labeling the nodes by the respective triads of integers m_1, m_2, m_3, as shown in Fig. 17.22, one starts by finding the intercepts of Π with the directions of the characteristic vectors. Letting such intercepts be (m_1^*, m_2^*, m_3^*), the triad $(r\, m_1^*, r\, m_2^*, r\, m_3^*)$ with $r \neq 0$ an integer, spans a set of planes parallel to Π. If M is the largest divisor of m_1^*, m_2^*, m_3^*, then the new triad $m_i' = m_i^*/M$ identifies the plane Π' parallel to Π and closest to the origin. Then, the inverse of the triad's elements are taken: $1/m_1'$, $1/m_2'$, $1/m_3'$. This avoids the occurrence of infinities; in fact, if Π were parallel to one of the characteristic vectors, say, $\mathbf{a}_i$, then m_i^* and m_i' would become infinite. One the other hand, using the inverse indices may bring to fractional numbers, a circumstance that must be avoided as well; so, as the last step, the new elements $1/m_i'$ are multiplied by the least multiple N of the m_i' that are not infinite:

$$(m_1'', m_2'', m_3'') = \left(\frac{N}{m_1'}, \frac{N}{m_2'}, \frac{N}{m_3'}\right). \tag{17.155}$$

The elements m_i'' thus found are the *Miller indices* of Π. They are enclosed in parentheses as in (17.155). By way of example, if $m_1^* = \infty$, $m_2^* = 2$, $m_3^* = 4$, then $M = 2$ so that $m_1' = \infty$, $m_2' = 1$, $m_3' = 2$. Calculating the inverse indices yields $1/m_1' = 0$, $1/m_2' = 1$, $1/m_3' = 1/2$; the least multiple is $N = 2$, so that the Miller indices are found to be $(0, 2, 1)$.

The indices that turn out to be negative are marked with a bar; for instance, in $(h\bar{k}l)$ the second index is negative. Some planes have the same symmetry; in cubic crystals this happens, for instance, to planes (100), (010), (001), $(\bar{1}00)$, $(0\bar{1}0)$, and $(00\bar{1})$. A set of planes with the same symmetry is indicated with braces, e.g., {100}. Examples of the (111), (001), and (010) planes are given in Fig. 17.23. As remarked in Sect. 24.4 about the silicon-oxidation process, the (111) plane has the highest concentration of atoms, followed by the {100} planes.

Symbols using three integers are also used to identify the crystal directions. To distinguish them from the symbols introduced so far, such triads of integers are

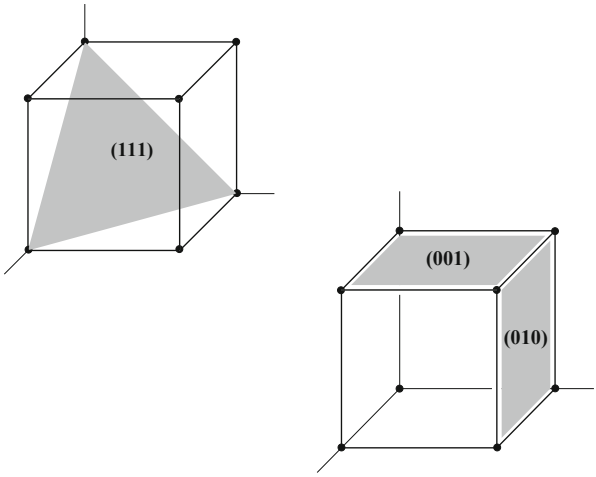

Fig. 17.23 Schematic representation of the (111) plane (*top left*) and of the (001) and (010) planes (*bottom right*) in a cubic crystal

enclosed in brackets. Consider, for instance, the line connecting nodes P and Q, oriented from P to Q. Letting m_{1P}, m_{2P}, m_{3P} be the coordinates of P, and the like for those of Q, one forms the new triad $m'_1 = m_{1Q} - m_{1P}$, $m'_2 = m_{2Q} - m_{2P}$, and $m'_3 = m_{2Q} - m_{2P}$. Then, the indices of the crystal direction are obtained as

$$[m''_1, m''_2, m''_3] = \left[\frac{m'_1}{M}, \frac{m'_2}{M}, \frac{m'_3}{M}\right], \tag{17.156}$$

with M the largest divisor of m'_1, m'_2, m'_3. Also in this case, negative indices are marked with a bar. By way of examples, the characteristic vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ in Fig. 17.22 are aligned, respectively, with the [100], [010], and [001] directions.

17.8.2 Examples of Translation Operators

A one-dimensional example of translation operator is easily found by considering the Taylor expansion of a function f around some position x:

$$f(x+l) = \sum_{n=0}^{\infty} \frac{l^n}{n!} \left(\frac{\mathrm{d}^n f}{\mathrm{d}x^n}\right)_{l=0} = \sum_{n=0}^{\infty} \frac{l^n}{n!} \frac{\mathrm{d}^n}{\mathrm{d}x^n} f(x) = \exp\left(l \frac{\mathrm{d}}{\mathrm{d}x}\right) f(x), \tag{17.157}$$

where the expression on the right stems from a formal application of the Taylor expansion of the exponential function, in which a numerical factor within the exponent is replaced with the operator $\mathrm{d}/\mathrm{d}x$. Extending the above reasoning to three dimensions yields

$$\mathcal{T}(\mathbf{l}) = \exp(\mathbf{l} \cdot \mathrm{grad}). \tag{17.158}$$

17.8.3 Symmetries of the Hamiltonian Operator

Given an operator $\mathcal{R}$, a second operator $\mathcal{A}$ is associated to $\mathcal{R}$ in the following manner [58, Sect. 1.5]:

$$\mathcal{A} f(\mathbf{r}) = f(\mathcal{R}^\dagger \mathbf{r}) \tag{17.159}$$

for all functions f. Thus, the action of $\mathcal{A}$ on f at $\mathbf{r}$ is the same as calculating the original function at $\mathbf{r}' = \mathcal{R}^\dagger \mathbf{r}$. Let $\mathcal{R}$ be unitary (Sect. 8.6.2), whence $\mathbf{r} = \mathcal{R}\mathbf{r}'$. A unitary operator acting on $\mathbf{r}$ leaves the norm $r = |\mathbf{r}|$ unchanged; as a consequence, the unitary operations possible on the coordinates are only those that perform a rotation or a reflexion of the coordinate axes, or both. It follows that the unit volume $\mathrm{d}\tau = \mathrm{d}^3 r$ is also invariant: $\mathrm{d}^3 \mathcal{R} r = \mathrm{d}^3 r$, this showing that $\mathcal{A}$ is unitary as well:

$$\int_\tau |\mathcal{A} f(\mathbf{r})|^2 \, \mathrm{d}^3 r = \int_{\tau'} |f(\mathbf{r}')|^2 \, \mathrm{d}^3 \mathcal{R} r' = \int_{\tau'} |f(\mathbf{r}')|^2 \, \mathrm{d}^3 r', \tag{17.160}$$

where τ' is the transformed domain. This reasoning does not apply to the translation operators $\mathcal{T}$. In fact, the operation $\mathcal{T}\mathbf{r} = \mathbf{r} + \mathbf{l}$ does not leave the norm of $\mathbf{r}$ unchanged. This shows in passing that $\mathcal{T}$ is not unitary. Other consequences of the above definitions and of the proof that $\mathcal{A}$ is unitary are

$$\mathcal{A} f(\mathcal{R}\mathbf{r}') = f(\mathbf{r}'), \qquad \mathcal{A}^\dagger f(\mathbf{r}') = f(\mathcal{R}\mathbf{r}'). \tag{17.161}$$

Also, $\mathcal{A}$ commutes with the operators that are invariant under the transformation $\mathbf{r} \leftarrow \mathcal{R}^\dagger \mathbf{r}$. In fact, if $\mathcal{B}$ is such an operator,

$$\mathcal{A}\mathcal{B}(\mathbf{r}) f(\mathbf{r}) = \mathcal{B}(\mathcal{R}^\dagger \mathbf{r}) f(\mathcal{R}^\dagger \mathbf{r}) = \mathcal{B}(\mathbf{r}) \mathcal{A} f(\mathbf{r}) \tag{17.162}$$

for all functions f. As $\mathcal{R}^\dagger$ is the inverse of $\mathcal{R}$, then $\mathcal{B}$ is also invariant under the transformation $\mathbf{r} \leftarrow \mathcal{R}\mathbf{r}$. As a consequence, $\mathcal{B}$ commutes also with $\mathcal{A}^\dagger$.

Let $\mathcal{B} v_n = b_n v_n$ be the eigenvalue equation for $\mathcal{B}$ (a discrete spectrum is assumed for the sake of simplicity). If b_n is s-fold degenerate, and $v_n^{(1)}, v_n^{(2)}, \ldots, v_n^{(s)}$ are s linearly-independent eigenfunctions corresponding to b_n, then

$$\mathcal{B} \sum_{i=1}^{s} c_i \, v_n^{(i)} = \sum_{i=1}^{s} c_i \, \mathcal{B} v_n^{(i)} = \sum_{i=1}^{s} c_i \, b_n \, v_n^{(i)} = b_n \sum_{i=1}^{s} c_i \, v_n^{(i)}, \tag{17.163}$$

namely, any non-vanishing linear combination of the form $\varphi_n = \sum_{i=1}^{s} c_i \, v_n^{(i)}$ is also an eigenfunction of $\mathcal{B}$ belonging to b_n. Let M be the space of all linear combinations of the form of φ_n; from (17.163) it follows that all members of M are eigenfunctions of $\mathcal{B}$ belonging to b_n. Conversely, all eigenfunctions of $\mathcal{B}$ belonging to b_n are members of M: letting $q_n \neq 0$ be one such eigenfunction, if q_n were not a member of M it would be $q_n - \sum_{i=1}^{s} c_i \, v_n^{(i)} \neq 0$ for all choices of the coefficients c_i. But this would imply that $q_n, v_n^{(1)}, v_n^{(2)}, \ldots, v_n^{(s)}$ are $s + 1$ linearly-independent eigenfunctions of b_n,

thus contradicting the hypothesis that the latter's degeneracy is of order s. Finally, if $\mathcal{A}$ commutes with $\mathcal{B}$ it is

$$\mathcal{B}\mathcal{A}\varphi_n = \mathcal{A}\mathcal{B}\varphi_n = \mathcal{A}b_n\varphi_n = b_n\mathcal{A}\varphi_n, \tag{17.164}$$

namely, $\mathcal{A}\varphi_n$ belongs to M.

In crystals, the unitary coordinate transformations $\mathbf{r}' = \mathcal{R}\mathbf{r}$ that leave the Hamiltonian operator $\mathcal{H}$ invariant are of particular interest. In fact, such coordinate transformations provide a method to study the degenerate eigenvalues of $\mathcal{H}$.

Let $\mathcal{B} = \mathcal{H}$, and let $\mathcal{H}$ be invariant under a coordinate transformation $\mathcal{R}\mathbf{r}$. If, in addition, $\mathcal{H}$ is translationally invariant and the periodic boundary conditions apply (Sect. 17.5.3), then the eigenfunctions w of $\mathcal{H}$ are Bloch functions, namely, they fulfill the Bloch theorem

$$w_i(\mathbf{r}+\mathbf{l},\mathbf{k}) = \exp(\mathrm{i}\,\mathbf{k}\cdot\mathbf{l})\, w_i(\mathbf{r},\mathbf{k}), \tag{17.165}$$

with $\mathbf{l}$ a translation vector and i the band index. Let $\mathcal{A}$ be the operator associated to $\mathcal{R}$. Then, from $\mathcal{H}\mathcal{A}^\dagger = \mathcal{A}^\dagger\mathcal{H}$,

$$\mathcal{H}\mathcal{A}^\dagger w_i(\mathbf{r},\mathbf{k}) = E_i(\mathbf{k})\,\mathcal{A}^\dagger w_i(\mathbf{r},\mathbf{k}), \tag{17.166}$$

with $E_i(\mathbf{k})$ the eigenvalue. One infers from (17.166) that, if $w_i(\mathbf{r},\mathbf{k})$ and $\mathcal{A}^\dagger w_i(\mathbf{r},\mathbf{k})$ are linearly independent, then the eigenvalue is degenerate. Such a degeneracy does not depend on the detailed form of the Hamiltonian operator, but only on its symmetry properties. For this reason, the degeneracy is called *essential*. If further degeneracies exist, that depend on the detailed form of $\mathcal{H}$, they are called *accidental*.

Let $M(\mathbf{k})$ be the space made of the linearly-independent eigenfunctions of $E(\mathbf{k})$, and of any non-vanishing linear combination of them, and define

$$v_i(\mathbf{r},\mathbf{k}') = \mathcal{A}^\dagger w_i(\mathbf{r},\mathbf{k}) = w_i(\mathcal{R}\mathbf{r},\mathbf{k}), \tag{17.167}$$

where symbol $\mathbf{k}'$ accounts for a possible influence on $\mathbf{k}$ of the coordinate transformation $\mathcal{R}\mathbf{r}$. Being an eigenfunction of $\mathcal{H}$, $v_i(\mathbf{r},\mathbf{k}')$ is a Bloch function,

$$v_i(\mathbf{r}+\mathbf{l},\mathbf{k}') = \exp(\mathrm{i}\,\mathbf{k}'\cdot\mathbf{l})\, v_i(\mathbf{r},\mathbf{k}'), \tag{17.168}$$

where $v_i(\mathbf{r}+\mathbf{l},\mathbf{k}') = w_i(\mathcal{R}\mathbf{r}+\mathcal{R}\mathbf{l},\mathbf{k})$. On the other hand, Bloch's theorem applied to $w_i(\mathcal{R}\mathbf{r}+\mathcal{R}\mathbf{l},\mathbf{k})$ yields

$$w_i(\mathcal{R}\mathbf{r}+\mathcal{R}\mathbf{l},\mathbf{k}) = \exp(j\mathbf{k}\cdot\mathcal{R}\mathbf{l})\, w_i(\mathcal{R}\mathbf{r},\mathbf{k}), \tag{17.169}$$

where the equality $\mathbf{k}\cdot\mathcal{R}\mathbf{l} = \mathcal{R}^\dagger\mathbf{k}\cdot\mathbf{l}$ holds due to the definition of adjoint operator. Comparison with the expression of the Bloch theorem applied to $v_i(\mathbf{r}+\mathbf{l},\mathbf{k}')$ provides $\mathbf{k}' = \mathcal{R}^\dagger\mathbf{k}$, whence

$$w_i(\mathcal{R}\mathbf{r},\mathbf{k}) = v_i(\mathbf{r},\mathcal{R}^\dagger\mathbf{k}). \tag{17.170}$$

In conclusion, if $w_i(\mathbf{r},\mathbf{k})$ is a Bloch function belonging to $M(\mathbf{k})$ and $\mathcal{R}\mathbf{r}$ a coordinate transformation that leaves the Hamiltonian operator invariant, then the eigenfunction

obtained by such a transformation also belongs to $M(\mathbf{k})$ and is labeled by $\mathcal{R}^\dagger\mathbf{k}$. The following also holds true,

$$\mathcal{H}v_i(\mathbf{r},\mathcal{R}^\dagger\mathbf{k}) = E_i(\mathcal{R}^\dagger\mathbf{k})\,v_i(\mathbf{r},\mathcal{R}^\dagger\mathbf{k}) \tag{17.171}$$

which, compared with $\mathcal{H}w_i(\mathbf{r},\mathbf{k}) = E_i(\mathbf{k})\,w_i(\mathbf{r},\mathbf{k})$, shows that

$$E_i(\mathcal{R}^\dagger\mathbf{k}) = E_i(\mathbf{k}). \tag{17.172}$$

The theory of this section is applied by way of example to the Hamiltonian operator of a system of K electrons and N nuclei, interacting through electrostatic forces, that was introduced in Sect. 16.2. The potential energy is (compare with (16.5))

$$U_e(\mathbf{r}) + U_a(\mathbf{r}) + U_{ea}(\mathbf{r},\mathbf{R}) + U_{\text{ext}}(\mathbf{r},\mathbf{R}), \tag{17.173}$$

with

$$U_e(\mathbf{r}) = \sum_{i,j=1}^{K} \frac{q^2}{4\,\pi\,\varepsilon_0\,|\mathbf{r}_i - \mathbf{r}_j|}, \qquad j \neq i. \tag{17.174}$$

Similar expressions hold for U_a and U_{ea} (the second relation in (16.1) and (16.2), respectively). If $U_{\text{ext}} = 0$, the potential energy is invariant upon the reflexion transformation $\mathcal{R}\mathbf{r} = -\mathbf{r}$, $\mathcal{R}\mathbf{R} = -\mathbf{R}$. Clearly, the kinetic part of the Hamiltonian operator is also invariant. In the adiabatic approximation (Sect. 16.3), the coordinates of the nuclei are fixed to the equilibrium positions $\mathbf{R}_0$, which preserves the reflexion invariance. Finally, the reflexion invariance is still preserved in the Hartree and Hartree–Fock approximations (Sects. 16.4 and 16.5, respectively), which also provide single-electron Hamiltonian operators that are translationally invariant. Due to lattice periodicity, the eigenfunctions of the Hamiltonian operator are Bloch functions. Denoting now with $\mathbf{r}$ the coordinates associated to a single electron, the transformation $\mathcal{R}\mathbf{r} = -\mathbf{r}$ corresponds to $\mathcal{R}^\dagger\mathbf{k} = -\mathbf{k}$ whence, from (17.172),

$$E_i(-\mathbf{k}) = E_i(\mathbf{k}). \tag{17.175}$$

This type of degeneracy is accidental because it depends on the detailed form of the Hamiltonian operator. If the crystal has also a reflection symmetry, then the reflexion invariance of the single-electron Hamiltonian operators occurs irrespective of the form of the interactions. In this case, the degeneracy is essential.

17.8.4 Kronig–Penney Model

The general method for solving the Schrödinger equation in a periodic lattice, shown in Sect. 17.6, is applied here to a one-dimensional case, where the potential energy is described as the series of equal barriers shown in Fig. 17.24. The approach is

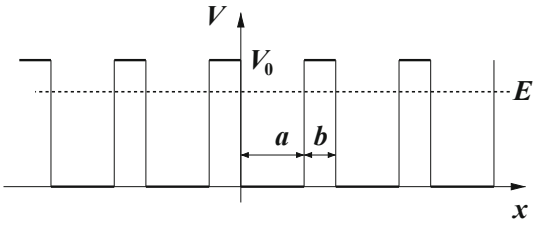

Fig. 17.24 Potential energy in the Kronig–Penney model

called *Kronig–Penney model*; it is amenable to an analytical solution and, despite its simplicity, is able to capture the main properties of the dispersion relation $E(\mathbf{k})$.

As shown in the figure, the potential energy is prescribed as $V = 0$ for $n\,(a+b) < x < n\,(a+b) + a$, and $V = V_0 > 0$ for $n\,(a+b) - b < x < n\,(a+b)$, with $n = 0, \pm 1, \pm 2 \ldots$ There is only one characteristic vector in the direct lattice, $\mathbf{a}_1 = (a+b)\,\mathbf{i}_1$; the corresponding characteristic vector of the reciprocal lattice is

$$\mathbf{b}_1 = \frac{\mathbf{i}_1}{a+b}. \tag{17.176}$$

As a consequence, the first Brillouin zone extends from $-\pi/(a+b)$ to $+\pi/(a+b)$ in the $\mathbf{i}_1$ direction. From the general properties of the time-independent Schrödinger equation (Sect. 8.2.3) it follows $E \geq 0$. As shown in Fig. 17.24, the case $0 < E < V_0$ is considered. A non-localized wave function w is expected even in the $E < V_0$ case due to the tunnel effect. From the Bloch theorem, the wave function has the form

$$w_k = u_k \, \exp(\mathrm{i}\, k\, x), \qquad u_k(x + a + b) = u_k(x), \tag{17.177}$$

where k belongs to the first Brillouin zone. In the intervals where $V = 0$ the Schrödinger equation reads

$$-w'' = \alpha^2\, w, \qquad \alpha = \sqrt{2\,m\,E}/\hbar > 0. \tag{17.178}$$

Replacing (17.177) into (17.178) yields

$$u_k'' + 2\,\mathrm{i}\,k\,u_k' - (k^2 - \alpha^2)\,u_k = 0, \tag{17.179}$$

whose associate algebraic equation has the roots

$$s = -\mathrm{i}\,k \pm \sqrt{-k^2 + (k^2 - \alpha^2)} = -\mathrm{i}\,k \pm \mathrm{i}\,\alpha. \tag{17.180}$$

The solution of (17.178) then reads

$$u_k^+ = c_1 \, \exp[\mathrm{i}\,(\alpha - k)\,x] + c_2 \, \exp[-\mathrm{i}\,(\alpha + k)\,x], \tag{17.181}$$

with c_1, c_2 undetermined coefficients. The procedure is similar in the intervals where $V = V_0$, and yields

$$w'' = \beta^2\, w, \quad \beta = \sqrt{2\,m\,(V_0 - E)}/\hbar, \quad u_k'' + 2\,\mathrm{i}\,k\,u_k' - (k^2 + \beta^2)\,u_k = 0, \tag{17.182}$$

$$s = -\mathrm{i}\,k \pm \sqrt{-k^2 + (k^2 + \beta^2)} = -\mathrm{i}\,k \pm \beta, \tag{17.183}$$

whence

$$u_k^- = c_3\, \exp\left[(\beta - \mathrm{i}\,k)\,x\right] + c_4\, \exp\left[-(\beta + \mathrm{i}\,k)\,x\right], \tag{17.184}$$

with c_3, c_4 undetermined coefficients. The regional solutions u_k^+, u_k^- must fulfill the continuity conditions imposed by the general properties of the Schrödinger equation; in addition, they must fulfill the periodicity condition prescribed by the Bloch theorem (second relation in (17.177)). To proceed, one focuses on the period $-b \le x \le a$, so that the continuity conditions at $x = 0$ for the function, $u_k^+(0) = u_k^-(0)$, and first derivative, $(u_k^+)'(0) = (u_k^-)'(0)$, provide

$$c_1 + c_2 = c_3 + c_4, \qquad \mathrm{i}\,\alpha\,(c_1 - c_2) = \beta\,(c_3 - c_4). \tag{17.185}$$

Combining (17.185),

$$c_1 = \sigma\, c_3 + \sigma^*\, c_4, \qquad c_2 = \sigma^*\, c_3 + \sigma\, c_4, \qquad 2\,\sigma = 1 - \mathrm{i}\,\beta/\alpha. \tag{17.186}$$

In turn, from the periodicity of u, namely, $u_k^+(a) = u_k^-(-b)$, and of u', namely, $(u_k^+)'(a) = (u_k^-)'(-b)$, one finds

$$c_1\, A + \frac{c_2}{A} = K\, L\, \left(\frac{c_3}{B} + c_4\, B\right), \qquad c_1\, A - \frac{c_2}{A} = -K\, L\, \left(\frac{c_3}{B} - c_4\, B\right) \mathrm{i}\, \frac{\beta}{\alpha}, \tag{17.187}$$

with

$$A = \exp\left(\mathrm{i}\,\alpha\, a\right), \quad B = \exp\left(\beta\, b\right), \quad K = \exp\left(\mathrm{i}\,k\, a\right), \quad L = \exp\left(\mathrm{i}\,k\, b\right). \tag{17.188}$$

Combining (17.187),

$$c_1 = \frac{KL}{A}\, \left(\frac{\sigma}{B}\, c_3 + \sigma^*\, B\, c_4\right), \qquad c_2 = A\, K\, L\, \left(\frac{\sigma^*}{B}\, c_3 + \sigma\, B\, c_4\right). \tag{17.189}$$

Eliminating c_1, c_2 between (17.186) and (17.189), finally provides an algebraic system in the two unknowns $c3, c4$:

$$\sigma\, \left(1 - \frac{KL}{AB}\right)\, c_3 + \sigma^*\, \left(1 - \frac{BKL}{A}\right)\, c_4 = 0, \tag{17.190}$$

$$\sigma^*\, \left(1 - \frac{AKL}{B}\right)\, c_3 + \sigma\, \left(1 - ABKL\right)\, c_4 = 0. \tag{17.191}$$

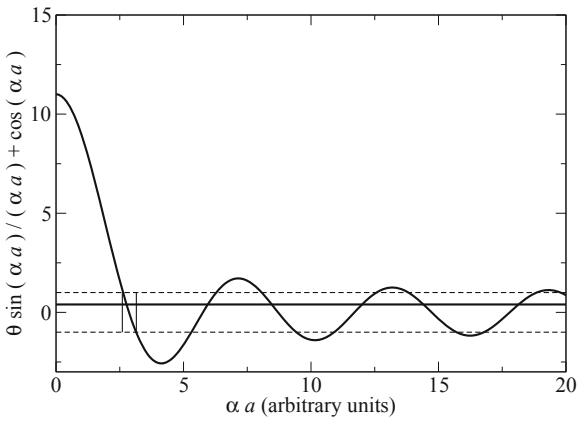

Fig. 17.25 Graphic solution of (17.194), with $\vartheta = 10$. The two *vertical lines* mark the values of $\alpha\, a$ delimiting the lowest band

As expected, the system is homogeneous, so a solution is possible only if the determinant vanishes. This in turn determines a relation between $\alpha(E)$, $\beta(E)$, and k, that eventually provides the dispersion relation $E(k)$. The determinant vanishes if

$$(\sigma^{*2} - \sigma^2)\left(KL + \frac{1}{KL}\right) = \sigma^{*2}\left(\frac{A}{B} + \frac{B}{A}\right) - \sigma^2\left(AB + \frac{1}{AB}\right). \qquad (17.192)$$

Introducing the expressions (17.186, 17.188) of σ, A, B, K, L transforms (17.192) into

$$\frac{\beta^2 - \alpha^2}{2\,\alpha\,\beta}\,\sin(\alpha\, a)\,\sinh(\beta\, b) + \cos(\alpha\, a)\,\cosh(\beta\, b) = \cos[k\,(a+b)], \qquad (17.193)$$

which has the form $F(E) = G(k)$. From this, the relation $E = E(k)$ can be determined. Note that $G(-k) = G(k)$ and $G[k + 2\,\pi/(a+b)] = G(k)$. As a consequence, the function $E(k)$ is even and has the periodicity of the reciprocal scaled lattice, as should be.

To the purpose of completing the analysis one may simplify (17.193) by considering a limiting case, namely, $V_0 \gg E$ so that, from (17.178, 17.182), the limit $\beta^2 \gg \alpha^2$ would result. This, however, would eliminate the tunnel effect and reduce the problem to that of a series of boxes. To avoid this outcome, the proper limiting case is $b \to 0$ and $V_0 \to \infty$, in such a way as to leave the area $b\,V_0$ of each barrier unchanged.[22] In other terms, one lets $b = \text{const}/V_0$, so that $\beta^2\, b \to \text{const} \neq 0$ while $\beta\, b \to 0$. It follows $\sinh(\beta\, b) \to \beta\, b$, $\cosh(\beta\, b) \to 1$ so that, letting $\vartheta = \lim(a\, b\, \beta^2/2)$, the $F(E) = G(k)$ relation (17.193) simplifies to

$$\vartheta\,\frac{\sin(\alpha\, a)}{\alpha\, a} + \cos(\alpha\, a) = \cos(k\, a), \qquad \vartheta > 0, \qquad \alpha = \frac{\sqrt{2mE}}{\hbar}. \qquad (17.194)$$

[22] The same type of limit is applicable to the single-barrier case, whose transmission coefficient is given in (11.22).

The function $E = E(k)$ can be determined by inverting (17.194); alternatively, it may be obtained in graphic form as shown in Fig. 17.25, where ϑ has been fixed to 10: given k, the right hand side of (17.194) is fixed at some value $-1 \leq \cos(k\,a) \leq 1$. The energy E is then found by seeking $\alpha\,a$ such that the two sides become equal. The horizontal, dashed lines in the figure correspond to $\cos(ka) = 1$ and $\cos(k\,a) = -1$; they limit the interval where (17.194) has real solutions. The horizontal, continuous line corresponds to $\cos(k\,a) = 0.4$, while the oscillating curve represents the left hand side of (17.194). The latter intercepts the $\cos(k\,a) = 0.4$ line at infinite points $\alpha_1\,a, \alpha_2\,a, \ldots$; from each α_i thus found, one determines the energy corresponding to the given k from the relation $\alpha_i = \sqrt{2mE_i}/\hbar$. Each branch of the multi-valued function $E(k)$ is then found by repeating the procedure for all values of k within the first Brillouin zone, this making $\cos(k\,a)$ to range from -1 to 1. In the figure, the two vertical lines mark the values of $\alpha\,a$ delimiting the lowest band. The following are also worth noting:

- Letting λ indicate the left hand side of (17.194), there are no real solutions for $\lambda > 1$ or $\lambda < -1$; the intervals with no real solutions are the forbidden bands. In fact, the k solutions in the forbidden bands are complex: it is $k\,a = \pm\mathrm{i}\,\log(\lambda + \sqrt{\lambda^2 - 1})$ when $\lambda > 1$, and $k\,a = \pi \pm \mathrm{i}\,\log(|\lambda| + \sqrt{\lambda^2 - 1})$ when $\lambda < -1$.
- At large energies the (17.194) relation tends to $\cos(\alpha a) = \cos(ka)$, namely, to the free-particle one: $k = \alpha = \sqrt{2mE}/\hbar$.
- Like in the general case, for a finite structure where the periodic boundary conditions are applied, the above calculation still holds, with k a discrete variable.

17.8.5 Linear, Monatomic Chain

The calculation of vibrational spectra has been carried out in general form in Sect. 17.7. Simple examples of application, with reference to a one-dimensional lattice, are given in this section and in the next one. Like the Kronig–Penney model used in Sect. 17.8.4 for determining the dispersion relation of electrons, the one-dimensional models of the lattice vibrations are amenable to analytical solutions; the latter, as shown below, are able to provide the explicit expression of the dispersion relation.

To begin, consider a one-dimensional monatomic lattice made of N_c cells (*linear, monatomic chain*). Let the lattice be aligned with the x axis, and the corresponding characteristic vector be $\mathbf{a} = a\,\mathbf{i}$, $a > 0$, with $\mathbf{i}$ the unit vector of the x axis. Finally, let the positions of the $N_c + 1$ nodes be $0, a, 2\,a, \ldots, n\,a, \ldots$. The translation vector associated to the nth node is $\mathbf{l}_n = n\,a\,\mathbf{i}$. Finally, it is assumed that the motion of each atom is constrained to the x axis, and the periodic boundary conditions are applied.

Due to the periodic boundary conditions the nodes of indices $n = 0$ and $n = N_c$ are actually the same node. As a one-dimensional case is considered, with $N_b = 1$, the total number of atoms is $N = N_c$. The number of the lattice's degrees of freedom is N_c, and the correspondence with the indices used in the general theory (compare

with (17.124)) is

$$m, n = 1, \dots, N_c, \qquad \alpha, \beta = 1, \qquad u, w = 1. \tag{17.195}$$

As only one atom per cell is present, one may assume that the equilibrium position of each nucleus coincides with that of a node. In the harmonic approximation the force acting on the rth nucleus is a linear function of the displacements:

$$F_r = -\sum_{k=1}^{N_c} c_{rk}\, h_k, \tag{17.196}$$

where all coefficients c_{rk} in general differ from 0. In real crystals, however, the interaction between nuclei becomes rapidly negligible as the distance increases. As a consequence, the dynamics of a nucleus may be tackled in a simplified manner by considering only the interaction with the neighboring nuclei to be effective. This is equivalent to letting $c_{rk} = 0$ when $|r - k| > 1$, whence

$$F_r = -c_r^{r-1}\, h_{r-1} - c_r^r\, h_r - c_r^{r+1}\, h_{r+1} = F_r(h_{r-1}, h_r, h_{r+1}). \tag{17.197}$$

In the coefficients of (17.197), the lower index refers to the node being acted upon by the force at the left hand side, the upper index refers to the node whose displacement contributes to such a force. When the nuclei of indices $r - 1$, r, $r + 1$ are in the equilibrium positions it is $F_r(0, 0, 0) = 0$ for all r. On the other hand, it is also $F_r(\delta, \delta, \delta) = 0$, with $\delta \neq 0$ an arbitrary displacement. In fact, when all displacements are equal, the interatomic distance remains the same as in the equilibrium condition. From $F_r(\delta, \delta, \delta) = 0$ it follows that the coefficients are connected by the relation

$$c_r^{r-1} + c_r^r + c_r^{r+1} = 0. \tag{17.198}$$

Moreover, on account of the fact that all atoms are identical and all equilibrium distances are also identical, it is $F_r(-\delta, 0, \delta) = 0$, $\delta \neq 0$, whence

$$c_r^{r-1} = c_r^{r+1}. \tag{17.199}$$

From (17.198, 17.199) it follows

$$c_r^r = -c_r^{r-1} - c_r^{r+1} = -2\, c_r^{r-1} = -2\, c_r^{r+1}. \tag{17.200}$$

Finally, the relation $F_r(0, \delta, 0) = -c_r^r\, \delta$, on account of the fact that $(0, 0, 0)$ is an equilibrium condition, shows that $c_r^r > 0$. As shown in Sect. 17.7 for the general case, due to the translational invariance the elastic coefficients do not depend on the cell index, but on the difference between cell indices (compare with (17.125)); in conclusion, letting

$$\chi = -c_r^{r-1} = -c_r^{r+1} > 0, \qquad c_r^r = 2\,\chi, \tag{17.201}$$

and letting μ be the common mass of the nuclei, the dynamics of the rth nucleus is described by the equation

$$\mu \ddot{h}_r = -\chi \ (2\, h_r - h_{r+1} - h_{r-1})\,. \tag{17.202}$$

The general theory shows that the displacement has the form

$$h_r = h_0 \ \exp(\mathrm{i}\, q\, r\, a - \mathrm{i}\, \omega\, t)\,, \tag{17.203}$$

(compare with (17.152)), where h_0 is a complex constant, $q\, r\, a = \mathbf{q} \cdot \mathbf{l}_r = q\, \mathbf{i} \cdot r\, a\, \mathbf{i}$, and $\omega = \omega(q)$. Replacing (17.203) in (17.202) and dividing by h_r yields

$$\mu\, \omega^2 = \chi \ \left[2 - \exp(\mathrm{i}\, q\, a) - \exp(-\mathrm{i}\, qa)\right] = 4\, \chi \ \sin^2(qa/2). \tag{17.204}$$

Defining $\tilde{\omega} = \sqrt{\chi/\mu}$ and remembering that ω is non negative, one finds the dispersion relation

$$\omega(q) = 2\tilde{\omega}\, |\sin(qa/2)|. \tag{17.205}$$

From the periodic boundary condition $h(r = N_c) = h(r = 0)$ one finds, with ν an integer,

$$\exp(\mathrm{i}\, q\, N_c\, a) = 1, \qquad q\, N_c\, a = 2\, \pi\, \nu, \qquad q = \frac{\nu}{N_c}\, \frac{2\, \pi}{a}. \tag{17.206}$$

Replacing this form of q within (17.203) and (17.205) shows that using $\nu + N_c$ instead of ν leaves h_r and ω unchanged. As a consequence, it is sufficient to consider only N_c consecutive values of ν, say, $\nu = 0, 1, \ldots, N_c - 1$, which in turn limit the possible values of q to an interval of length $2\, \pi/a$. This was expected, because the values of the indices in (17.195) are such that the number of eigenvalues of the problem is N_c. Thus, the dispersion relation has only one branch, given by (17.205). One also notes that $2\, \pi/a$ is the size of the first Brillouin zone in the one-dimensional case. Typically, the interval of q is made to coincide with the first Brillouin zone, namely, $-\pi/a \leq q < +\pi/a$. Also, as mentioned in Sect. 17.7, in most cases q is treated as a continuous variable. The phase and group velocities are

$$u_f = \frac{\omega}{q} = \pm a\, \tilde{\omega}\, \frac{\sin(q\, a/2)}{qa/2}, \qquad u = \frac{\mathrm{d}\omega}{\mathrm{d}q} = \pm a\, \tilde{\omega}\, \cos(qa/2), \tag{17.207}$$

respectively, where the positive (negative) sign holds when q is positive (negative). At the boundary of the Brillouin zone it is $qa/2 = \pi/2$, whence $\omega = 2\, \tilde{\omega}$, $u_f = \pm a\tilde{\omega}/\pi$, $u = 0$. Near the center of the Brillouin zone it is $\omega \simeq a\tilde{\omega}\, |q|$, $u_f \simeq u \simeq \pm a\tilde{\omega}$. At the center it is $\omega = 0$.

The dispersion relation (17.205) normalized to $\tilde{\omega} = \sqrt{\chi/\mu}$ is shown in Fig. 17.26 as a function of $qa/2$. The range of the first Brillouin zone is $-\pi/2 \leq qa/2 \leq +\pi/2$. Remembering that the wavelength corresponding to q is $\lambda = 2\, \pi/q$, the interval near the origin, where the phase and group velocities are equal to each other and independent of q, corresponds to the largest values of the vibrations' wavelength. As some of these wavelengths fall in the audible range, the branch is called *acoustic branch*.

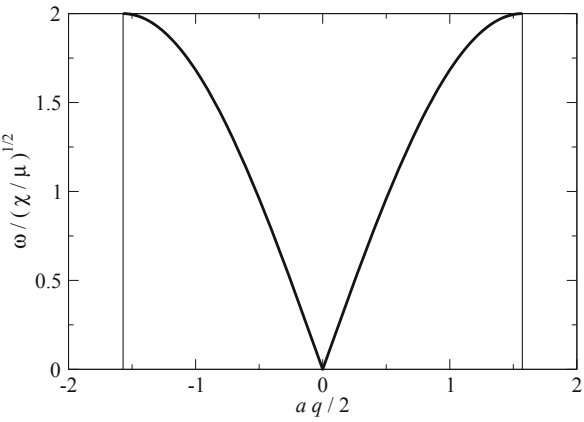

Fig. 17.26 Normalized dispersion relation of a linear, monatomic chain. The *vertical lines*, placed at $a\,q/2 = \pm\pi/2$, are the limits of the first Brillouin zone

17.8.6 Linear, Diatomic Chain

As a second example, consider a one-dimensional lattice made of N_c cells, with a two-atom basis. Let the lattice be aligned with the x axis, and the corresponding characteristic vector be $\mathbf{a} = a\,\mathbf{i}$, $a > 0$, with $\mathbf{i}$ the unit vector of the x axis. Finally, let the positions of the $N_c + 1$ nodes be $0, a, 2\,a, \ldots, n\,a, \ldots$. The translation vector associated to the nth node is $\mathbf{l}_n = n\,a\,\mathbf{i}$. Finally, it is assumed that the motion of each atom is constrained to the x axis, and the periodic boundary conditions are applied.

Due to the periodic boundary conditions the nodes of indices $n = 0$ and $n = N_c$ are actually the same node. As a one-dimensional case is considered, with $N_b = 2$, the total number of atoms is $N = 2\,N_c$. The number of degrees of freedom of the lattice is $2\,N_c$, and the correspondence with the indices (17.124) used in the general theory of Sect. 17.7 is

$$m, n = 1, \ldots, N_c, \qquad \alpha, \beta = 1, 2, \qquad u, w = 1. \tag{17.208}$$

As two atoms per cell are present, one may assume that the equilibrium position of one type of nucleus coincides with that of a node. Such nuclei will be given the index $\alpha, \beta = 1$, while the other nuclei will be given the index $\alpha, \beta = 2$. In the harmonic approximation the force acting on a nucleus is a linear function of the displacements:

$$F_r = -\sum_{k=1}^{2\,N_c} c_{rk}\,h_k. \tag{17.209}$$

In real crystals, the interaction between nuclei becomes rapidly negligible as the distance increases. Following the same reasoning as in Sect. 17.8.5, the dynamics of a nucleus is tackled in a simplified manner by considering only the interaction with the neighboring nuclei to be effective. This is equivalent to letting $c_{rk} = 0$ when $|r - k| > 1$. For a nucleus of type 1 the neighboring nuclei are of type 2. It is assumed

that the node numbering is such, that the neighbors of interest belong to the cells of indices $r-1$ and r. It follows

$$F_{r,1} = -c_{r,1}^{r-1,2}\, h_{r-1,2} - c_{r,1}^{r,1}\, h_{r,1} - c_{r,1}^{r,2}\, h_{r,2} = F_{r,1}(h_{r-1,2}, h_{r,1}, h_{r,2}). \tag{17.210}$$

In the coefficients of (17.210), the left-lower index refers to the node being acted upon by the force at the left hand side, the left-upper index refers to the node whose displacement contributes to such a force, the right-lower and right-upper indices refer to the nucleus type. When the nuclei involved are in the equilibrium positions it is $F_{r,1}(0,0,0) = 0$. On the other hand, it is also $F_{r,1}(\delta,\delta,\delta) = 0$, with $\delta \neq 0$ an arbitrary displacement. In fact, when all displacements are equal the interatomic distance remains the same as in the equilibrium condition. From $F_{r,1}(\delta,\delta,\delta) = 0$ it follows

$$c_{r,1}^{r-1,2} + c_{r,1}^{r,1} + c_{r,1}^{r,2} = 0. \tag{17.211}$$

As on the other hand there is no special symmetry in the interaction of the nucleus of indices $r, 1$ with the neighboring ones, it is in general (in contrast to the case of a monatomic linear chain) $c_{r,1}^{r-1,2} \neq c_{r,1}^{r,2}$. Finally, the relation $F_{r,1}(0,\delta,0) = -c_{r,1}^{r,1}\,\delta$, on account of the fact that $(0,0,0)$ is an equilibrium condition, shows that $c_{r,1}^{r,1} > 0$.

The calculation is then repeated for a nucleus of type 2, whose neighboring nuclei are of type 1. Due to the node numbering chosen here, the neighbors of interest belong to the cells of indices r and $r+1$. As a consequence,

$$F_{r,2} = -c_{r,2}^{r,1}\, h_{r,1} - c_{r,2}^{r,2}\, h_{r,2} - c_{r,2}^{r+1,1}\, h_{r+1,1} = F_{r,2}(h_{r,1}, h_{r,2}, h_{r+1,1}). \tag{17.212}$$

By the same reasoning leading to (17.211) one finds

$$c_{r,2}^{r,1} + c_{r,2}^{r,2} + c_{r,2}^{r+1,1} = 0, \tag{17.213}$$

where, like in the case of (17.211), it is in general $c_{r,2}^{r,1} \neq c_{r,2}^{r+1,1}$. Finally, the relation $F_{r,2}(0,\delta,0) = -c_{r,2}^{r,2}\,\delta$, on account of the fact that $(0,0,0)$ is an equilibrium condition, shows that $c_{r,2}^{r,2} > 0$. Due to the lattice periodicity the coefficients do not depend on the cell index, whence one lets

$$-\chi_1 = c_{r-1,1}^{r-1,2} = c_{r,1}^{r,2} = c_{r+1,1}^{r+1,2} = \ldots \qquad -\chi_2 = c_{r,1}^{r-1,2} = c_{r+1,1}^{r,2} = c_{r+2,1}^{r+1,2} = \ldots \tag{17.214}$$

Remembering the invariance relation $c_{m\alpha u}^{n\beta w} = c_{n\beta w}^{m\alpha u}$ of the general theory one also finds

$$c_{r,2}^{r,1} = c_{r,1}^{r,2} = -\chi_1, \qquad c_{r,2}^{r+1,1} = c_{r+1,1}^{r,2} = c_{r,1}^{r-1,2} = -\chi_2, \tag{17.215}$$

whence

$$c_{r,1}^{r,1} = c_{r,2}^{r,2} = \chi_1 + \chi_2 > 0. \tag{17.216}$$

Finally, from the relations $F_{r,1}(0,0,\delta) = \chi_1\,\delta$, $F_{r,1}(\delta,0,0) = \chi_2\,\delta$, it follows that $\chi_1 > 0$, $\chi_2 > 0$, on account of the fact that $(0,0,0)$ is an equilibrium condition. Letting μ_1, μ_2 be the masses of the two types of nuclei, the dynamics of the rth nuclei is described by the equations

$$\mu_1\,\ddot{h}_{r,1} = -\chi_1\,\left(h_{r,1} - h_{r,2}\right) - \chi_2\,\left(h_{r,1} - h_{r-1,2}\right), \tag{17.217}$$

$$\mu_2\,\ddot{h}_{r,2} = -\chi_1\,\left(h_{r,2} - h_{r,1}\right) - \chi_2\,\left(h_{r,2} - h_{r+1,1}\right), \tag{17.218}$$

where the displacements have the form

$$h_{r,1(2)} = h_{0,1(2)}\,\exp\left(\mathrm{i}\,q r a - \mathrm{i}\,\omega t\right), \tag{17.219}$$

thanks to the general theory. In (17.219), $h_{0,1}$, $h_{0,2}$ are complex constants, $q\,r\,a = \mathbf{q}\cdot\mathbf{l}_r = q\,\mathbf{i}\cdot r\,a\,\mathbf{i}$, and $\omega = \omega(q)$. Replacing (17.219) in (17.217, 17.218) and dividing by $h_{r,1}$, $h_{r,2}$, respectively, yields

$$\mu_1\,\omega^2\,h_{0,1} = \chi_1\,\left(h_{0,1} - h_{0,2}\right) + \chi_2\,\left[h_{0,1} - h_{0,2}\,\exp\left(-\mathrm{i}\,qa\right)\right], \tag{17.220}$$

$$\mu_2\,\omega^2\,h_{0,2} = \chi_1\,\left(h_{0,2} - h_{0,1}\right) + \chi_2\,\left[h_{0,2} - h_{0,1}\,\exp\left(+\mathrm{i}\,qa\right)\right]. \tag{17.221}$$

Defining A_{11}, A_{12}, A_{21}, A_{22} such that

$$\mu_1\,A_{11} = \chi_1 + \chi_2, \qquad -\mu_1\,A_{12} = \chi_1 + \chi_2\,\exp\left(-\mathrm{i}\,qa\right), \tag{17.222}$$

$$\mu_2\,A_{22} = \chi_1 + \chi_2, \qquad -\mu_2\,A_{21} = \chi_1 + \chi_2\,\exp\left(+\mathrm{i}\,qa\right), \tag{17.223}$$

the homogeneous algebraic system (17.220, 17.221) transforms into

$$\left(A_{11} - \omega^2\right)\,h_{0,1} + A_{12}\,h_{0,2} = 0, \qquad A_{21}\,h_{0,1} + \left(A_{22} - \omega^2\right)\,h_{0,2} = 0. \tag{17.224}$$

The trace $T = A_{11} + A_{22}$ and determinant $D = A_{11}\,A_{22} - A_{12}\,A_{21}$ of the matrix formed by A_{11}, A_{12}, A_{21}, A_{22} read

$$T = \frac{\mu_1 + \mu_2}{\mu_1\mu_2}\,\left(\chi_1 + \chi_2\right), \qquad D = 2\,\frac{\chi_1\,\chi_2}{\mu_1\,\mu_2}\,\left[1 - \cos\left(qa\right)\right]. \tag{17.225}$$

The eigenvalues ω^2 are found by solving the algebraic equation $(\omega^2)^2 - T\,\omega^2 + D = 0$, whose discriminant is

$$\Delta(q) = T^2 - 4\,D = \left[\frac{\left(\chi_1 + \chi_2\right)\left(\mu_1 + \mu_2\right)}{\mu_1\,\mu_2}\right]^2 + 8\,\frac{\chi_1\,\chi_2}{\mu_1\,\mu_2}\,\left[\cos\left(qa\right) - 1\right]. \tag{17.226}$$

Remembering that $\chi_1, \chi_2 > 0$, the minimum Δ_m of (17.226) occurs for $q = \pm\pi/2$. Letting $K_\chi = (\chi_1 - \chi_2)^2/(4\,\chi_1\,\chi_2) \geq 0$ and $K_\mu = (\mu_1 - \mu_2)^2/(4\,\mu_1\,\mu_2) \geq 0$, one finds the relation

$$\Delta(q) \geq \Delta_m = 16\,\frac{\chi_1\,\chi_2}{\mu_1\,\mu_2}\,\left[\left(1 + K_\chi\right)\left(1 + K_\mu\right) - 1\right] \geq 0, \tag{17.227}$$

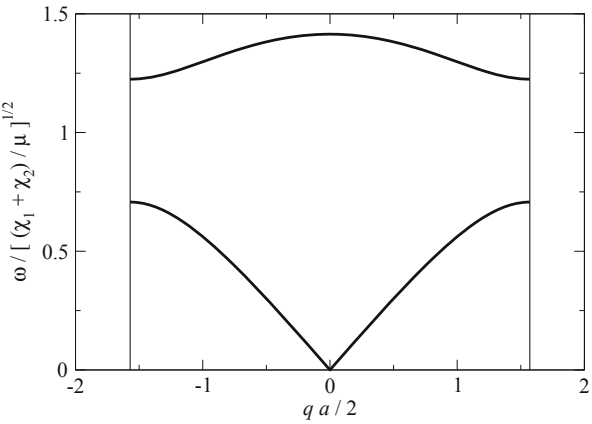

Fig. 17.27 Normalized dispersion relation of a linear, diatomic chain with $\mu_1 = \mu_2 = \mu$ and $\chi_1 = 3\,\chi_2$. The *vertical lines*, placed at $a\,q/2 = \pm\pi/2$, are the limits of the first Brillouin zone

showing that the discriminant is non negative. It follows that the eigenvalues ω^2 are real, as should be. The solution of the algebraic equation provides two branches of the dispersion relation, to be found by taking the square root of

$$\omega^2 = \frac{T}{2} \pm \frac{1}{2}\sqrt{\Delta(q)}. \tag{17.228}$$

Observing that $\Delta(0) = T^2$, one finds that selecting the minus sign in (17.228) provides the branch that contains $\omega = 0$. As in the case of the monatomic chain, this branch is called *acoustic branch*. In the other branch it is always $\omega > 0$; in ionic crystals like, e.g., sodium chloride, the frequencies typical of this branch are excited by infrared radiation. For this reason, the branch is called *optical branch*.

The acoustic and optical branch of a linear, diatomic chain are shown in Fig. 17.27, where $\mu_1 = \mu_2 = \mu$ is assumed for simplicity. Letting $\tilde{\omega} = \sqrt{(\chi_1 + \chi_2)/\mu}$, one finds for the acoustic branch

$$\frac{\omega_{\rm ac}^2}{\tilde{\omega}^2} = 1 - \left[1 - 4\,\frac{\chi_1\,\chi_2}{(\chi_1 + \chi_2)^2}\,\sin^2\left(\frac{qa}{2}\right)\right]^{1/2}. \tag{17.229}$$

At the center of the first Brillouin zone it is $\omega_{\rm ac} = 0$, while the maximum of $\omega_{\rm ac}$ is reached at the boundary $qa/2 = \pm\pi/2$ of the zone. For the optical branch one finds

$$\frac{\omega_{\rm op}^2}{\tilde{\omega}^2} = 1 + \left[1 - 4\,\frac{\chi_1\,\chi_2}{(\chi_1 + \chi_2)^2}\,\sin^2\left(\frac{qa}{2}\right)\right]^{1/2}. \tag{17.230}$$

At the center of the first Brillouin zone $\omega_{\rm op}$ reaches its maximum. The minimum of $\omega_{\rm op}$ is reached at the boundary $qa/2 = \pm\pi/2$ of the zone. The discretization of q due to the periodic boundary conditions, and the definitions of the phase and group velocity, are the same as those already given for the monatomic lattice.

From (17.229, 17.230), the distance $G = \omega_{\rm op}(q = \pm\pi/a) - \omega_{\rm ac}(q = \pm\pi/a)$ between the minimum of the optical branch and the maximum of the acoustic branch

fulfills the relation

$$\frac{G}{\tilde{\omega}} = \left(1 + \frac{|\chi_1 - \chi_2|}{\chi_1 + \chi_2}\right)^{1/2} - \left(1 - \frac{|\chi_1 - \chi_2|}{\chi_1 + \chi_2}\right)^{1/2}. \tag{17.231}$$

It is $G > 0$ if $\chi_2 \neq \chi_1$, whereas $G = 0$ if $\chi_2 = \chi_1$. The latter case is called *degenerate*.

17.8.7 Analogies

It is interesting to note the analogy between the expression of the energy of the electromagnetic field *in vacuo*, described as a superposition of modes (Eqs. (5.38, 5.40)), and that of a system of vibrating nuclei (Eqs. (3.48) and (17.118)). In essence, the two expressions derive from the fact the in both cases the energy is a positive-definite, symmetric quadratic form. In the case of the electromagnetic field the form is exact because of the linearity of the Maxwell equations; for the vibrating nuclei the form is approximate because of the neglect of the anharmonic terms (Sect. 3.13.1).

Other analogies exist between the dispersion relation $E(\mathbf{k})$ of the electrons subjected to a periodic potential energy, worked out in Sect. 17.6, and the dispersion relation $\omega(\mathbf{q})$, worked out in Sect. (17.7). Both relations are even and periodic in the reciprocal, scaled lattice; both have a branch structure, the difference being that the number of branches of $\omega(\mathbf{q})$ is finite because the number of degrees of freedom of the vibrating lattice is finite, whereas that of $E(\mathbf{k})$ is infinite.