CAMBRIDGE
Cambridge University Press
978-0-521-51460-6 - Understanding Modern Transistors and Diodes
David L. Pulfrey
Excerpt
More information

1 Introduction

It is highly probable that you will use a laptop computer when doing the exercises in this book. If so, you may be interested to know that the central processing unit of your computer resides in a thin sliver of silicon, about 1 square centimetre in area. This small chip contains over 100,000,000 Si MOSFETs,[1] each about a thousand times smaller than the diameter of a human hair! The slender computer that you nonchalantly stuff into your backpack has more computing power than the vacuum-tube computers that occupied an entire room when I was a student over 40 years ago.

When you are reading this book, you may be distracted by an incoming call on your cell 'phone. That may get you wondering what's inside your sleek 'mobile'. If you opened it up, and knew where to look, you'd find some GaAs HBTs.[2] These transistors can operate at the high frequencies required for local-area-network telecommunications, and they can deliver the power necessary for the transmission of signals.

Of course, a cell 'phone nowadays is no longer just a replacement for those clunking, tethered, hand-sets of not so long ago: it is also a camera and a juke box. The immense storage requirements of these applications are met by Flash memory, comprising more millions of Si MOSFETs.

Your cell 'phone is really a PDA,[3] and probably also allows internet access, in which case you may wonder how signals from around the globe find their way into your machine. Somewhere in the communications chain there's probably a low-noise amplifier to receive tiny signals and not add undue noise to them. GaAs HBTs are good for this, but even better are InP HEMTs.[4] If satellites are involved, then the base station will employ high-power transistors, possibly lateral-diffused Si MOSFETs, or maybe GaN HJFETs.[5]

So, without straying very far from where you are sitting as you read this, you have tangible evidence of the dramatic influence electronics has on the way many of us conduct our business and recreation. All the different transistors mentioned above are described in this book, and are grouped according to their ability to perform: in high-speed digital logic; at high frequencies; with low noise; at high output power; in semiconductor memory.

---

[1] Metal-Oxide-Semiconductor Field-Effect Transistors.
[2] Heterojunction Bipolar Transistors.
[3] Personal Digital Assistant.
[4] High Electron Mobility Transistors.
[5] Heterojunction Field-Effect Transistors.

Cambridge University Press
978-0-521-51460-6 - Understanding Modern Transistors and Diodes
David L. Pulfrey
Excerpt
More information

Of course, our electronics-oriented activities would not be possible if the supply of electricity were curtailed. This could happen, either by the exhaustion of the Earth's store of fossil fuels, or by the threat to our habitable environment that the extraction and use of them entails. Alternate, and renewable, forms of electrical energy generation are desirable; photovoltaics, using semiconductor diodes as solar cells, is an attractive proposition. How solar cells work is described in this book. We look at traditional Si cells, and at both thin-film cells and tandem cells for possible implementation in the future.

You may know that about 20% of the world's energy consumption goes into producing light. Glance up at the incandescent light bulb that is illuminating your room: it's so inefficient that if you had a few of them in use, then you probably wouldn't need to heat your study in winter! Again, some alternative is needed; LEDs[6] using diodes made from compound semiconductors are beginning to make an impact in this area. We describe how high-brightness LEDs work, and look at ways of producing white light.

To understand the operation of all these transistors and diodes, and to provide the knowledge base that will enable you to understand new devices as they appear, and to design better devices yourself, a solid, physical understanding of semiconductors must be attained. The first part of this book is devoted to this. The emphasis is on Quantum Mechanics, as this branch of physics is needed increasingly to understand transistors as they move from the micro- to the nano-realm, and also, of course, to understand interactions between electrons and holes and photons in optoelectronic diodes.

The book ends with a brief look at cylindrical nanotransistors, the future development of which may perhaps involve you?

Enjoy the book!

---

[6] Light-Emitting Diodes.

CAMBRIDGE

Cambridge University Press
978-0-521-51460-6 - Understanding Modern Transistors and Diodes
David L. Pulfrey
Excerpt
More information

2 Energy band basics

Louis de Broglie, in his Ph.D. thesis of 1924, postulated that every object that has momentum p also has a wavelength λ:

$$p = \frac{h}{\lambda}, \tag{2.1}$$

where h is Planck's constant. Macroscopic objects of our everyday experience have extremely short wavelengths, so they are invariably viewed as particles, with a point mass and an observable trajectory. Contrarily, microscopic objects can have much longer wavelengths, and may do wave-like things, such as diffract around other microscopic objects. Electrons and atoms are microscopic objects, so when we need to consider them both together we must take a quantum-mechanical, rather than a classical, approach. This is what we do in this chapter. Our initial goal is to develop the concept of energy bands, representing ranges of permissible energies for electrons within a solid. We then seek to provide an understanding of related concepts that are used throughout this book: electron states, crystal momentum, band structure, holes, effective mass, energy band diagrams. These objectives are most directly arrived at from a consideration of the periodic nature of the potential through which the electrons would move in a perfectly crystalline material.

2.1 Periodic structures

Crystalline structures are based on a matrix of points called a **Bravais lattice**. For the Group IV semiconductors and most of the III-V semiconductors that are considered in this book, the Bravais lattice is the face-centred cubic lattice. To this underlying structure are added the actual atoms that constitute the **basis** of a particular material. The basis for Si, Ge, GaAs, InP, for example, comprises two atoms, which are shown as any neighbouring pair of shaded and unshaded atoms in Fig. 2.1. Each atom occupies a site on a face-centred cubic lattice, so the actual structure comprises two, interpenetrating, face-centred cubic lattices. When the two atoms are the same, as in the elemental semiconductors Si and Ge, the structure is called **diamond**. When the two atoms are different, e.g., Ga and As, the structure is referred to as **sphalerite** or **zinc blende**. The bonding of atoms in these structures is tetragonal, as shown by the linkages in Fig. 2.1.

Instead of trying to deal with the countless numbers of atoms that comprise an actual piece of crystalline material, it is often convenient to capture the structural essence of a

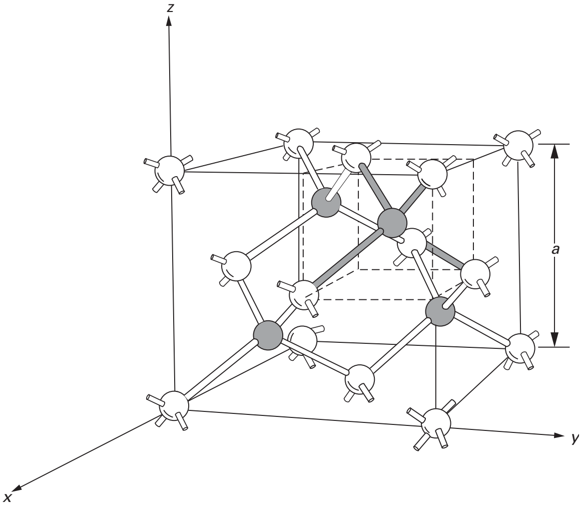

Figure 2.1 The diamond and sphalerite crystal structure. There are two, interpenetrating, face-centred cubic (FCC) lattices, one comprising the shaded atoms and the other comprising the unshaded atoms. The corresponding points in each FCC lattice are displaced by $\frac{a}{4}(\hat{x}+\hat{y}+\hat{z})$, where a, the **lattice constant**, is the length of the side of the cube. Adapted from Sze [1], © John Wiley & Sons, Inc. 1985, reproduced with permission.

crystal in its primitive unit cell, or, simply, **primitive cell**. This is a volume, containing precisely one lattice point, from which, by appropriate rotations and translations, the space of the Bravais lattice can be exactly filled. There is no unique primitive unit cell for a given Bravais lattice, and one of them is shown by the dashed lines in Fig. 2.1. Another primitive unit cell is the **Wigner-Seitz primitive cell**, the construction of which is illustrated in Fig. 2.2 for a simple face-centred rectangular matrix of unshaded atoms. The primitive unit cell in this case is a hexagon, which also contains one of the shaded atoms from an identical matrix of atoms. Thus, this particular crystal structure has a basis of two. For a real 3-D crystal the lines between nearest-neighbour atoms are bisected by planes; and for the face-centred cubic lattice the Wigner-Seitz cell is a rhombic dodecahedron [2, Fig. 1.8b].

2.2 Periodic potential

To illustrate the relationship between energy and momentum in a crystalline material, we consider a ‘toy’ structure comprising a one-dimensional array of primitive cells, with each cell having a basis of unity, and the atom being monovalent (see Fig. 2.3a). The potential energy of a single electron due to Coulombic interaction with the ion cores of the monovalent atoms is shown in Fig. 2.3b. However, we are not interested here in the precise form of the potential energy: we are only concerned with its periodicity. Therefore, we reduce the potential-energy profile to the delta-function representation shown in Fig. 2.3d. Don’t be alarmed that the last profile might not be very realistic:

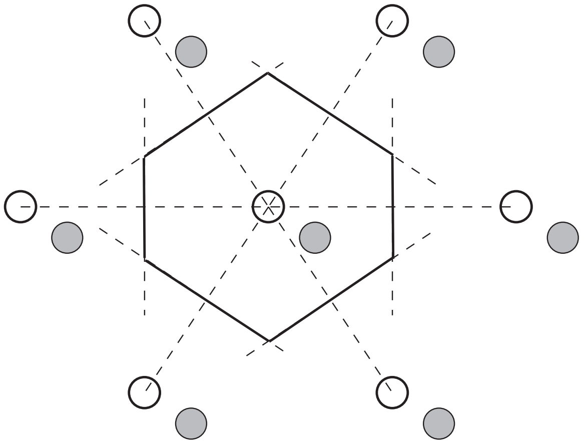

Figure 2.2 Example of a 2-D crystal comprising simple face-centred rectangular arrays of unshaded and shaded atoms. The Wigner-Seitz primitive unit cell is shown by the solid lines. These lines connect the perpendicular bisectors of the lines joining one unshaded atom to each neighbouring unshaded atom. One atom from the shaded array falls within the primitive unit cell; thus, this crystal structure has a basis of two atoms.

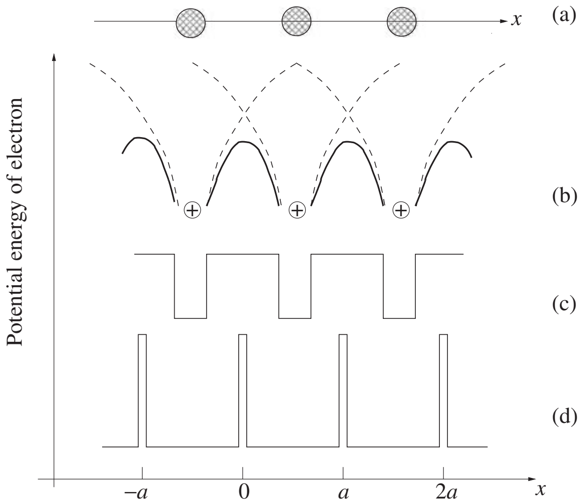

Figure 2.3 (a) 1-D periodic array of primitive cells, each cell containing one monovalent atom. (b) 1-D Coulombic potential energy for an electron in the 1-D array. Dashed lines are the potential energies due to a single ion core. Solid lines are the total potential energy. (c) 1-D square well representation of (b). (d) 1-D delta-function representation of (c).

even Fig. 2.3b is inaccurate, as it omits effects such as: the potential energy of an electron due to the proximity of other electrons; the different spacing between atoms in different directions of the real (3-D) crystal; and the possible presence of dissimilar elements in the crystal, e.g., as in compound semiconductors, such as GaAs. The important fact is that any periodic potential leads to the revelation of energy bands, and, therefore, will do

Cambridge University Press
978-0-521-51460-6 - Understanding Modern Transistors and Diodes
David L. Pulfrey
Excerpt
More information

for our present purpose.[1] The profile in Fig. 2.3d, comprising N delta-function potential barriers spaced a apart, can be expressed as

$$U(x) = \beta \sum_{l=0}^{N-1} \delta(x - la)\,, \tag{2.2}$$

where $\delta(x)$ is the Dirac delta function and β is some constant.[2]

2.3 Schrödinger's equation

When considering the fine details of an electron's motion in a solid, we need to consider its wave-like nature. The appropriate equation is the Schrödinger Wave Equation, which was originally postulated in 1925 to provide a formal description of the experimentally observed, discrete frequencies of light emission from an excited hydrogen atom. You can have confidence in the equation because, in the intervening 80+ years, no experiments have been reported that give results contrary to the predictions of the equation. The form of the equation of interest to us here is the **time-independent Schrödinger Wave Equation**, i.e., in one dimension,

$$-\frac{\hbar^2}{2m_0}\frac{d^2\psi(x)}{dx^2} + U(x)\psi(x) = E\psi(x)\,, \tag{2.3}$$

where m_0 is the electron rest mass, $\psi(x)$ is the position-dependent part of the **electron wavefunction** $\Psi(x,t)$, U is the potential energy and E is the total energy.[3]

Thinking in terms of conservation of energy, it can be seen that the first term in (2.3) must relate to kinetic energy. Often, the first two terms are grouped together and described as the **Hamiltonian** of the system

$$\mathcal{H}\psi = E\psi\,, \tag{2.4}$$

where the Hamiltonian $\mathcal{H}$ operates on the wavefunction to describe the total energy of the system.

Niels Bohr's statistical interpretation of the wavefunction is particularly helpful in getting a feeling for what Ψ really is: $\Psi\,\Psi^* dx \equiv |\Psi(x,t)|^2\,dx$ is the probability of finding the electron between x and $(x + dx)$ at time t.[4] If the electron is somewhere in x (1-D case), then it follows that $\int_{-\infty}^{+\infty} |\Psi(x,t)|^2\,dx = 1$. Equivalently, $\int_{-\infty}^{+\infty} |\psi(x)|^2\,dx = 1$. Thus, $\Psi(x,t)$ and $\psi(x)$ enable us to compute the probability of finding an electron

---

[1] If you insist on giving some physical significance to the potential profile in Fig. 2.3d, then you may wish to view the electron as being largely confined to the vicinity of an atom, but having some probability of tunnelling to a neighbouring, identical, region through a thin potential barrier.

[2] The property of the delta function that is relevant here is: $\delta(y) = 0$ if $y \neq 0$, and $\delta(y) = \infty$ if $y = 0$.

[3] This equation follows from the full, time-dependent Schrödinger Wave Equation, which describes the full wavefunction, i.e., in the 1-D case, $\Psi(x,t)$. In all our work we will take the potential energy to be independent of time. This allows the full equation to be solved by the method of Separation of Variables, for which solutions are simply: $\Psi(x,t) = \psi(x)\,f(t)$, where $f(t) = \exp(-iEt/\hbar)$ and $E = \hbar\omega$. Thus, we can solve (2.3) for $\psi(x)$, and then always multiply by $f(t)$ to get the full time dependence if we need it.

[4] The superscript * denotes the complex conjugate.

Cambridge University Press
978-0-521-51460-6 - Understanding Modern Transistors and Diodes
David L. Pulfrey
Excerpt
More information

somewhere in space at some time. This is how quantum mechanics works: it deals in probabilities. This is not an inadequacy of the theory; it is a description of how Nature appears to work at the level of very tiny entities.

2.4 Energy bands

Consider the periodic delta-function potential in Fig. 2.3d. Here, we use it to develop an understanding of energy bands, closely following the treatment of Griffiths [3]. In the region $0 < x < a$ the potential energy is zero, so, from (2.3)

$$\frac{d^2\psi}{dx^2} + g^2\psi(x) = 0\,, \tag{2.5}$$

where

$$g = \frac{\sqrt{2m_0E}}{\hbar}\,. \tag{2.6}$$

The general solution is

$$\psi(x) = A\sin(gx) + B\cos(gx), \qquad (0 < x < a)\,. \tag{2.7}$$

A and B are constants that need to be evaluated by considering the boundary conditions. The general rules are:

- ψ must be continuous at a boundary;
- $d\psi/dx$ must be continuous at a boundary, except when the potential energy goes to infinity.[5]

In our problem we have lots of boundaries, and at each one $U \to \infty$. Fortunately, because of the periodic nature of the potential, we can reach a solution quite easily by appealing to **Bloch's Theorem**, which states that for a periodic potential $U(x + a) = U(x)$, the solutions to Schrödinger's equation satisfy

$$\psi_k(x) = u_k(x)e^{ikx}\,, \tag{2.8}$$

where $u_k(x)$ has the periodicity of the lattice, and the subscript k indicates that $u(x)$ has different functional forms for different values of the **Bloch wavenumber** k. Note that if u is not periodic but is a constant, then the Bloch wave becomes a plane wave. Therefore, a Bloch wave, given by (2.8), is a plane wave modulated by a function that has the periodicity of the lattice. An alternative way of stating Bloch's Theorem follows from (2.8), namely

$$\psi_k(x + a) = e^{ika}\psi_k(x)\,. \tag{2.9}$$

---

[5] If there is a discontinuity in $d\psi/dx$, then the kinetic-energy term in (2.3) $\to \infty$, but the equation is still satisfied if $U \to \infty$. When we resort to the 'Effective-mass Schrödinger Wave Equation', the boundary condition for the derivative of ψ must also include what we shall call the effective mass, if this property changes across the boundary (see Section 2.11).

Note that this equation does not state that $\psi_k(x)$ is periodic, but it does lead to $|\psi_k(x)|^2$ being periodic. The latter is comforting because one would expect an electron to have an equal probability of being at any of the identical sites in the linear array. The periodicity breaks down at the edges of the crystal, but that shouldn't have a significant effect on the electrons deep within the crystal if the array is very long compared to the separation between atoms, i.e., if N, the number of primitive cells, is very large. Mathematically, we can impose complete periodicity by bending the array into a circle so that $x = -a$ follows $x = (N-2)a$ in Fig. 2.3d. We then have a convenient, so-called **periodic boundary condition**:

$$\psi_k(x + Na) = \psi_k(x)\,. \tag{2.10}$$

Using this in (2.9), yields

$$e^{ikNa}\psi_k(x) = \psi_k(x)\,, \tag{2.11}$$

from which it is clear that

$$k = \frac{2\pi n}{Na}\,, \qquad (n = 0, \pm 1, \pm 2, \pm 3, \cdots)\,, \tag{2.12}$$

where n is an integer. (2.9) can now be used to obtain the wavefunction in the region $-a < x < 0$ of Fig. 2.3d:

$$\psi_k(x) = e^{-ika}[A \sin g(x+a) + B \cos g(x+a)], \qquad (-a < x < 0)\,. \tag{2.13}$$

Now that we have expressions for the wavefunctions in adjoining regions we can use the matching conditions for ψ and $d\psi/dx$ to evaluate or eliminate the constants A and B. Matching the wavefunctions at $x = 0$ gives

$$B = e^{-ika}[A \sin(ga) + B \cos(ga)]\,. \tag{2.14}$$

Because of the delta function, the derivative of ψ is not continuous at $x = 0$, so we need to find the discontinuity in order to get another expression linking A and B. For $U(x) = \beta\delta(x)$, which comes from (2.2), the discontinuity is

$$\Delta(\frac{d\psi}{dx}) = \frac{2m_0\beta}{\hbar^2}\psi(0)\,.^6 \tag{2.15}$$

Thus, it follows from the derivatives of ψ at $x = 0$ that

$$gA - e^{-ika}g[A\cos(ga) - B\sin(ga)] = \frac{2m_0\beta}{\hbar^2}B\,. \tag{2.16}$$

---

[6] To obtain this, integrate Schrödinger's equation over a tiny interval spanning $x = 0$. The integral of the $d^2\psi/dx^2$ term is precisely the discontinuity we seek. It is equal to the integrals over the $E\psi$ and $U\psi$ terms. In the former term E is a constant and ψ is finite, so integrating over an infinitesimal interval gives zero. The same would usually be true for the $U\psi$ term, but because $U = \infty$ at $x = 0$, the integral is finite and equals $\beta\psi(0)$, where we have used another property of the delta function: $\int_{-\infty}^{\infty}\delta(x)\,dx = 1$.

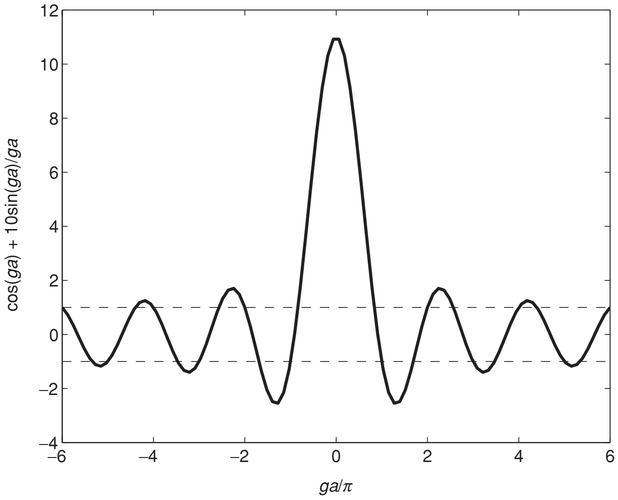

Figure 2.4 Plot of (2.17) for $\left[\frac{m_0\beta a}{\hbar^2}\right] = 10$, showing the allowed values of ga, i.e., those within the dashed lines. The forbidden values of ga lie in the areas outside the dashed lines.

From (2.14) and (2.16), after some manipulation, an expression devoid of A and B results:

$$\cos(ka) = \cos(ga) + \left[\frac{m_0\beta a}{\hbar^2}\right]\frac{\sin(ga)}{ga}\,. \tag{2.17}$$

This key equation unlocks the secret of bands: the right-hand side is a function of ga, and g is a function of the energy E from (2.6), but the left-hand side decrees that $f(ga)$ must be bounded by ±1. Thus, values of E are only allowed when $-1 \le f(ga) \le 1$. This is illustrated by the plot of (2.17) in Fig. 2.4. Note that this figure is arbitrarily truncated at $g = 6\pi/a$, but, in reality, g could be extended indefinitely; thus, there are an infinite number of ranges of allowed energy, each one of which is called an **energy band**.

The energy bands corresponding to the allowed values of ga, and the forbidden regions (bandgaps) separating the bands, are usually displayed on a plot of energy E versus Bloch wavevector k. The version shown in Fig. 2.5 is known as an **extended-zone** plot. The first zone spans the range $-\pi/a < k < \pi/a$; the second zone is split into two: $-2\pi/a < k < -\pi/a$ and $\pi/a < k < 2\pi/a$; etc. Thus each zone extends over a range of $2\pi/a$ in k. From (2.12), it is seen that the corresponding range in n is N, the number of primitive calls. As the latter number will be usually very large in semiconductor devices, the separation of neighbouring k values ($=2\pi/Na$), is so small that the E-k relation appears continuous within a band.

An E-k plot is often interpreted as an energy-momentum relationship. This is because, from (2.1), momentum can be written as $\hbar k$, where $\hbar = h/2\pi$ is Dirac's constant, and $k = 2\pi/\lambda$ is the general relationship between wavelength and wavevector. For the specific case of a Bloch wavevector, $\hbar k$ is called the **crystal momentum**. The crystal

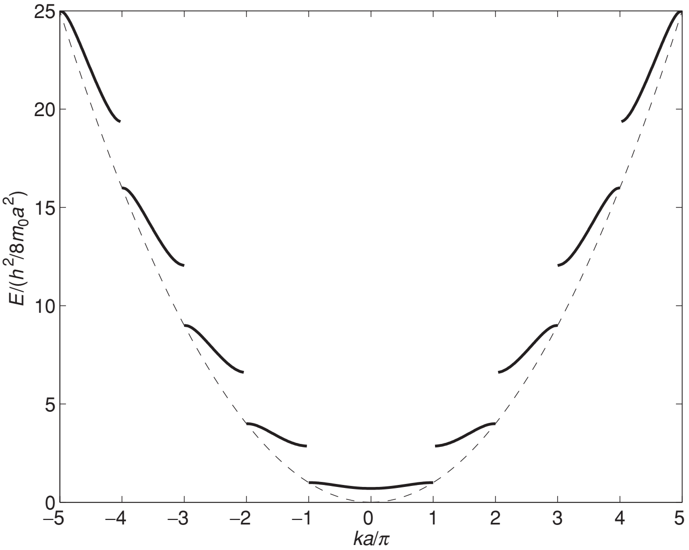

Figure 2.5 'Extended-zone' plot of energy (from Fig. 2.4 and Equation (2.6)) for the first five allowed energy bands. For example, the first band of ga runs from $ga = 0.83\pi$ to $ga = 1.00\pi$ (see Fig. 2.4). This range of ga values, and their negatives, are then used in (2.6) to obtain the first allowed band of energies. The corresponding ka range for the first band is $-\pi < ka < \pi$. The parabola shown by the dashed curve is the E-k relation for a free electron. Note how the allowed bands become closer to this parabola as the energy increases, indicating the increasing 'freedom' of the higher energy electrons.

momentum is not the actual mechanical momentum of the electron: it is the momentum of the electron due to the action of applied forces, as we show in Section 2.9.

2.5 Reduced-zone plot

An alternative way of displaying the E-k relationship is to compress all of its information into the first zone. This is achieved by horizontally shifting each of the curves from the higher order zones in the extended-zone plot by an appropriate multiple of $2\pi/a$. For example, consider the positive wavevectors in the 4th and 5th zones, i.e., $3\pi/a < k < 5\pi/a$. Now, write the wavevector as

$$k = \frac{4\pi}{a} + k' \,, \tag{2.18}$$

where the new wavevector k' is constrained to $-\pi/a \le k' \le \pi/a$, i.e., to the first zone. The Bloch wavefunction from (2.8) then becomes

$$\begin{aligned} \psi_k(x) &= u_k(x)e^{i4\pi x/a}e^{ik'x} \\ &\equiv u'_k(x)e^{ik'x} \\ &= \psi_{k'}(x) \,. \end{aligned} \tag{2.19}$$