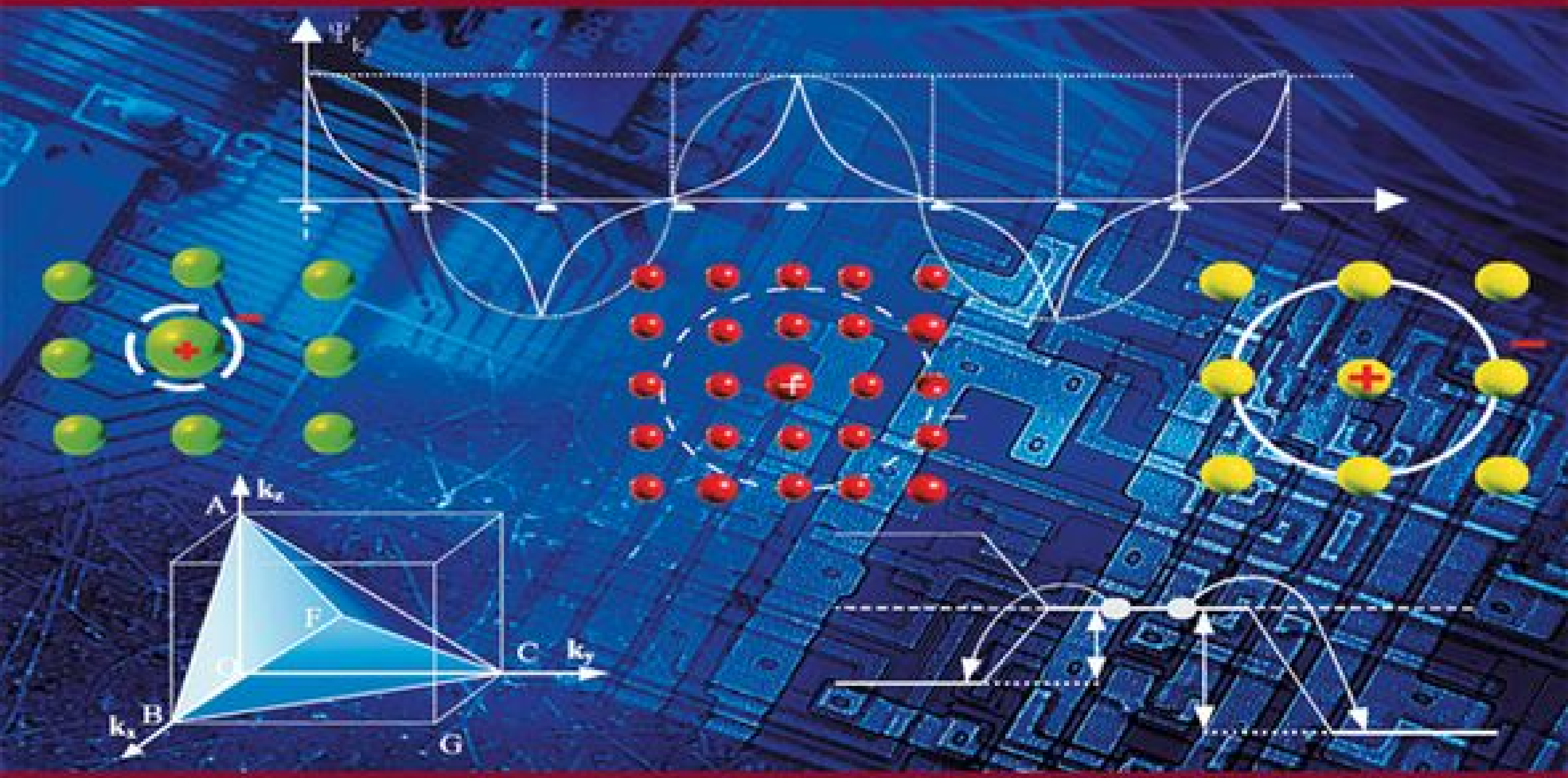


Solid-State Physics for Electronics

André Moliton

ISTE

WILEY

Solid-State Physics for Electronics

André Moliton

Series Editor
Pierre-Noël Favennec

ISTE

WILEY

Solid-State Physics for Electronics

Solid-State Physics for Electronics

André Moliton

Series Editor
Pierre-Noël Favennec

ISTE WILEY

First published in France in 2007 by Hermes Science/Lavoisier entitled*: Physique des matériaux pour l'électronique* © LAVOISIER, 2007
First published in Great Britain and the United States in 2009 by ISTE Ltd and John Wiley & Sons, Inc.

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

Library of Congress Cataloging-in-Publication Data

Moliton, André.
[Physique des matériaux pour l'électronique. English]
Solid-state physics for electronics / André Moliton.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-84821-062-2
1. Solid state physics. 2. Electronics--Materials. I. Title.
QC176.M5813 2009
530.4'1--dc22
2009016464

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN: 978-1-84821-062-2

Cover image created by Atelier Istatis.
Printed and bound in Great Britain by CPI Antony Rowe, Chippenham and Eastbourne.

Table of Contents

Foreword

A student that has attained a MSc degree in the physics of materials or electronics will have acquired an understanding of basic atomic physics and quantum mechanics. He or she will have a grounding in what is a vast realm: solid state theory and electronic properties of solids in particular. The aim of this book is to enable the step-by-step acquisition of the fundamentals, in particular the origin of the description of electronic energy bands. The reader is thus prepared for studying relaxation of electrons in bands and hence transport properties, or even coupling with radiance and thus optical properties, absorption and emission. The student is also equipped to use by him- or herself the classic works of taught solid state physics, for example, those of Kittel, and Ashcroft and Mermin.

This aim is reached by combining qualitative explanations with a detailed treatment of the mathematical arguments and techniques used. Valuably, in the final part the book looks at structures other than the macroscopic crystal, such as quantum wells, disordered materials, etc., towards more advanced problems including Peierls transition, Anderson localization and polarons. In this, the author's research specialization of conductors and conjugated polymers is discernable. There is no doubt that students will benefit from this well placed book that will be of continual use in their professional careers.

Michel SCHOTT

Emeritus Research Director (CNRS),
Ex-Director of the Groupe de Physique des Solides (GPS),
Pierre and Marie Curie University, Paris, France

Introduction

This volume proposes both course work and problems with detailed solutions. It is the result of many years' experience in teaching at MSc level in applied, materials and electronic physics. It is written with device physics and electronics students in mind. The book describes the fundamental physics of materials used in electronics. This thorough comprehension of the physical properties of materials enables an understanding of the technological processes used in the fabrication of electronic and photonic devices.

The first six chapters are essentially a basic course in the rudiments of solid-state physics and the description of electronic states and energy levels in the simplest of cases. The last four chapters give more advanced theories that have been developed to account for electronic and optical behaviors of ordered and disordered materials.

The book starts with a physical description of weak and strong electronic bonds in a lattice. The appearance of energy bands is then simplified by studying energy levels in rectangular potential wells that move closer to one another. Chapter 2 introduces the theory for free electrons where particular attention is paid to the relation between the nature of the physical solutions to the number of dimensions chosen for the system. Here, the important state density functions are also introduced. Chapter 3, covering semi-free electrons, is essentially given to the description of band theory for weak bonds based on the physical origin of permitted and forbidden bands. In Chapter 4, band theory is applied with respect to the electrical and electronic behaviors of the material in hand, be it insulator, semiconductor or metal. From this, superlattice structures and their application in optoelectronics is described. Chapter 5 focuses on ordered solid-state physics where direct lattices, reciprocal lattices, Brillouin zones and Fermi surfaces are good representations of electronic states and levels in a perfect solid. Chapter 6 applies these representations to metals and semiconductors using the archetypal examples of copper and silicon respectively. An excursion into the preparation of alloys is also proposed.

The last four chapters touch on theories which are rather more complex. Chapter 7 is dedicated to the description of the strong bond in 1D media. Floquet's theorem, which is a sort of physical analog for the Hückel's theorem that is so widely used in physical chemistry, is established. These results are extended to 3D media in Chapter 8, along with a simplified presentation of silicon band theory. The huge gap between the discovery of the working transistor (1947) and the rigorous establishment of silicon band theory around 20 years later is highlighted. Chapter 9 is given over to the description of energy levels in real solids where defaults can generate localized levels. Amorphous materials are well covered, for example, amorphous silicon is used in non-negligible applications such as photovoltaics. Finally, Chapter 10 contains a description of the principal *quasi*-particles in solid state, electronic and optical physics. Phonons are thus covered in detail. Phonons are widely used in thermics; however, the coupling of this with electronic charges is at the origin of phonons in covalent materials. These polarons, which often determine the electronic transport properties of a material, are described in all their possible configurations. Excitons are also described with respect to their degree of extension and their presence in different materials. Finally, the coupling of an electromagnetic wave with electrons or with (vibrating) ions in a diatomic lattice is studied to give a classical description of *quasi*-particles such as plasmons and polaritons.

Chapter 1

Introduction: Representations of Electron-Lattice Bonds

1.1. Introduction

This book studies the electrical and electronic behavior of semiconductors, insulators and metals with equal consideration. In metals, conduction electrons are naturally more numerous and freer than in a dielectric material, in the sense that they are less localized around a specific atom.

Starting with the dual wave-particle theory, the propagation of a de Broglie wave interacting with the outermost electrons of atoms of a solid is first studied. It is this that confers certain properties on solids, especially in terms of electronic and thermal transport. The most simple potential configuration will be laid out first (Chapter 2). This involves the so-called flat-bottomed well within which free electrons are simply thought of as being imprisoned by potential walls at the extremities of a solid. No account is taken of their interactions with the constituents of the solid. Taking into account the fine interactions of electrons with atoms situated at nodes in a lattice means realizing that the electrons are no more than semi-free, or rather "*quasi*-free", within a solid. Their bonding is classed as either "weak" or "strong" depending on the form and the intensity of the interaction of the electrons with the lattice. Using representations of weak and strong bonds in the following chapters, we will deduce the structure of the energy bands on which solid-state electronic physics is based.

1.2. Quantum mechanics: some basics

1.2.1. *The wave equation in solids: from Maxwell's to Schrödinger's equation via the de Broglie hypothesis*

In the theory of wave-particle duality, Louis de Broglie associated the wavelength (λ) with the mass (m) of a body, by making:

$$\lambda = \frac{h}{mv}. \tag{1.1}$$

For its part, the wave propagation equation for a vacuum (here the solid is thought of as electrons and ions swimming in a vacuum) is written as:

$$\Delta s - \frac{1}{c^2}\frac{\partial^2 s}{\partial t^2} = 0. \tag{1.2}$$

If the wave is monochromatic, as in:

$$s = A(x,y,z)e^{-i\omega t} = A(x,y,z)e^{-i2\pi\nu t}$$

then $\Delta s = \Delta A e^{-i\omega t}$ and $\frac{\partial^2 s}{\partial t^2} = -\omega^2 A e^{-i\omega t}$ (without modifying the result we can interchange a wave with form $s = A(x,y,z)e^{i\omega t} = A(x,y,z)e^{i2\pi\nu t}$). By introducing $\lambda = 2\pi\frac{c}{\omega}$ (length of a wave in a vacuum), wave propagation equation [1.2] can be written as:

$$\Delta A + \frac{\omega^2}{c^2}A = 0 \tag{1.3}$$

$$\rightleftarrows$$

$$\Delta A + \frac{4\pi^2}{\lambda^2}A = 0. \tag{1.3'}$$

A particle (an electron for example) with mass denoted m, placed into a time-independent potential energy $V(x, y, z)$, has an energy:

$$E = \frac{1}{2} m\mathrm{v}^2 + V$$

(in common with a wide number of texts on quantum mechanics and solid-state physics, this book will inaccurately call potential the “potential energy” – to be denoted V).

The speed of the particle is thus given by

$$\mathrm{v} = \sqrt{\frac{2(E - V)}{m}}. \qquad [1.4]$$

The de Broglie wave for a frequency $\nu = \frac{E}{h}$ can be represented by the function Ψ (which replaces the s in equation [1.2]):

$$\Psi = \psi e^{-i\,2\pi\nu t} = \psi e^{-2\pi i \frac{E}{h} t} = \psi e^{-i \frac{E}{\hbar} t} = \psi e^{-i\,\omega t}. \qquad [1.5]$$

Accepting with Schrödinger that the function ψ (amplitude of Ψ) can be used in an analogous way to that shown in equation [1.3’], we can use equations [1.1] and [1.4] with the wavelength written as:

$$\lambda = \frac{h}{m\mathrm{v}} = \frac{h}{\sqrt{2m(E - V)}}, \qquad [1.6]$$

so that:

$$\Delta\psi + \frac{2m}{\hbar^2}(E - V)\psi = 0. \qquad [1.7]$$

This is the Schrödinger equation that can be used with crystals (where V is periodic) to give well defined solutions for the energy of electrons. As we shall see, these solutions arise as permitted bands, otherwise termed valence and conduction bands, and forbidden bands (or "gaps" in semiconductors) by electronics specialists.

1.2.2. *Form of progressive and stationary wave functions for an electron with known energy (E)*

In general terms, the form (and a point defined by a vector $\vec{r}$) of a wave function for an electron of known energy (*E*) is given by:

$$\Psi(\vec{r},t) = \psi(\vec{r})e^{-j\omega t} = \psi(\vec{r})e^{-j\frac{E}{\hbar}t},$$

where $\psi(\vec{r})$ is the wave function at amplitudes which are in accordance with Schrödinger's equation [1.7]:

– if the resultant wave $\Psi(r,t)$ is a stationary wave, then $\psi(\vec{r})$ is real;

– if the resultant wave $\Psi(\vec{r},t)$ is progressive, then $\psi(\vec{r})$ takes on the form $\psi(\vec{r}) = f(\vec{r})e^{j\vec{k}.\vec{r}}$ where $f(\vec{r})$ is a real function, and $\vec{k} = \frac{2\pi}{\lambda}\vec{u}$ is the wave vector.

1.2.3. *Important properties of linear operators*

1.2.3.1. *If the two (linear) operators H and T are commutative, the proper functions of one can also be used as the proper functions of the other*

For the sake of simplicity, non-degenerate states are used. For a proper function ψ of *H* corresponding to the proper non-degenerate value (α), we find that:

$$H\psi = \alpha\psi$$

Multiplying the left-hand side of the equation by *T* gives:

$$TH\psi = T\alpha\psi = \alpha T\psi.$$

As $[H, T] = 0$, we can write:

$$HT\psi = \alpha T\psi.$$

This equation shows that $T\psi$ is a proper function of H with the proper value α. Hypothetically, this proper value is non-degenerate. Therefore, comparing the latter equation with the former ($H\psi = \alpha\psi$, indicating that ψ is a proper function of H for the same proper value α), we now find that $T\psi$ and ψ are collinear. This is written as:

$$T\psi = t\psi.$$

This equation in fact signifies that ψ is a proper function of T with the proper value being the coefficient of collinearity (t) (QED).

1.2.3.2. *If the operator H remains invariant when subject to a transformation using coordinates (T), then this operator H commutes with operator (T) associated with the transformation*

Here are the respective initial and final states (with initial on the left and final to the right):

energy: $E \xrightarrow{T} E'$

Hamiltonian: $H \xrightarrow{T} TH = H' = H$ (invariance of H under effect of T)

wave function: $\psi \xrightarrow{T} T\psi = \psi'$.

Similarly, the application of the operator T to the quantity $H\psi$, with H being invariant under T's effect, gives:

$$H\psi \xrightarrow{T} T(H\psi) = H'\psi' \overset{H'=H}{=} H\psi' \overset{\psi'=T\psi}{=} HT\psi.$$

We thus find:

$$TH\psi = HT\psi,$$

from which:

$$\left[H,T\right] = 0 \quad \text{QED.}$$

1.2.3.3. *The consequence*

If the operator H is invariant to the effect of the operator T, then the proper functions of T can be used as the proper functions of H.

1.3. Bonds in solids: a free electron as the zero order approximation for a weak bond; and strong bonds

1.3.1. *The free electron: approximation to the zero order*

The electric conduction properties of metals historically could have been derived from the most basic of theories, that of free electrons. This would assume that the conduction (or free) electrons move within a flat-bottomed potential well. In this model, the electrons are simply imprisoned in a potential well with walls that coincide with the limits of the solid. The potential is zero between the infinitely high walls. This problem is studied in detail in Chapter 2 with the introduction of the state density function that is commonly used in solid-state electronics. In three dimensions, the problem is treated as a potential box.

In order to take the electronic properties of semiconductors and insulators into account (where the electrons are no longer free), and indeed improve the understanding of metals, the use of more elaborate models is required. The finer interactions of electrons with nuclei situated at nodes throughout the solid are brought into play so that the well's flat bottom (where $V = V^0 = 0$) is perturbed or even strongly modified by the generated potentials. In a crystalline solid where the atoms are spread periodically in certain directions, the potential is also periodic and has a depth which depends on the nature of the solid.

Two approaches can be considered, depending on the nature of the bonds. If the well depth is small (weak bond) then a treatment of the initial problem (free electron) using perturbation theory is possible. If the wells are quite deep, for example as in a covalent crystal with electrons tied to given atoms through strong bonds, then a more global approach is required (using Hückels theories for chemists or Floquet's theories for physicists).

1.3.2. *Weak bonds*

This approach involves improving the potential box model. This is done by the electrons interacting with a periodical internal potential generated by a crystal lattice (of Coulumbic potential varying $1/r$ with respect to the ions placed at nodes of the lattice). In Figure 1.1, we can see atoms periodically spaced a distance "*a*" apart. Each of the atoms has a radius denoted "*R*" (Figure 1.1a). A 1D representation of the potential energy of the electrons is given in Figure 1.1b. The condition $a < 2R$ has been imposed.

Depending on the direction defined by the line *Ox* that joins the nuclei of the atoms, when an electron goes towards the nuclei, the potentials diverge. In fact, the study of the potential strictly in terms of *Ox* has no physical reality as the electrons here are conduction electrons in the external layers. According to the line (*D*) that does not traverse the nuclei, the electron-nuclei distance no longer reaches zero and potentials that tend towards finite values join together. In addition, the condition $a < 2R$ decreases the barrier that is midway between adjacent nuclei by giving rise to a strong overlapping of potential curves. This results in a solid with a periodic, slightly fluctuating potential. The first representation of the potential as a flat-bottomed bowl (zero order approximation for the electrons) is now replaced with a periodically varying bowl. As a first approximation, and in one dimension ($r \equiv x$), the potential can be described as:

$$V(x) = w_0 \cos \frac{2\pi}{a} x .$$

The term w_0, and the associated perturbation of the crystalline lattice, decrease in size as the relation $a < 2R$ becomes increasingly valid. In practical terms, the smaller "*a*" is with respect to $2R$, then the smaller the perturbation becomes, and the more justifiable the use of the perturbation method to treat the problem becomes. The corresponding approximation (first order approximation with the Hamiltonian perturbation being given by $H^{(1)} = w_0 \cos \frac{2\pi}{a} x$) is that of a semi-free electron and is an improvement over that for the free electron (which ignores $H^{(1)}$). The theory that results from this for the weak bond can equally be applied to the metallic bond, where there is an easily delocalized electron in a lattice with a low value of w_0 (see Chapter 3).

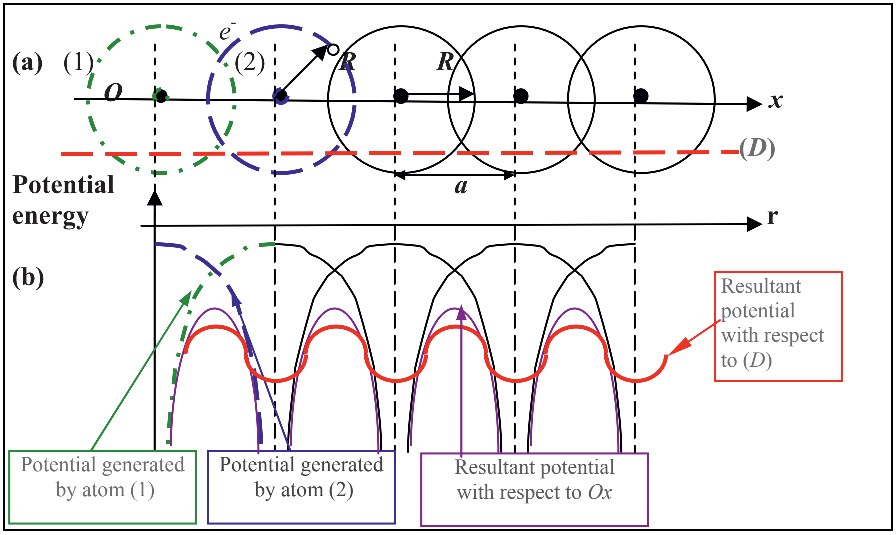


Figure 1.1. *Weak bonds and: (a) atomic orbitals (s orbitals with radius R) of a periodic lattice (period = a) respecting the condition a < 2R; (b) in 1D, the resultant potential energy (thick line) seen by electrons*

1.3.3. *Strong bonds*

The approach used here is more "chemical" in its nature. The properties of the solid are deduced from chemical bonding expressed as a linear combination of atomic orbitals of the constituent atoms. This reasoning is all the more acceptable when the electrons remain localized around a specific atom. This approximation of a strong bond is moreover justified when the condition $a \geq 2R$ is true (Figure 1.2a), and is generally used for covalent solids where valence electrons remain localized around the two atoms that they are bonding.

Once again, analysis of the potential curve drawn with respect to Ox gives a function which diverges as the distance between the electrons and the nuclei is reduced. With respect to the line D, this discontinuity of the valence electrons can be suppressed in two situations, namely (see also Figure 1.2b):

– If $a >> 2R$, then very deep potential wells appear, as there is no longer any real overlap between the generated potentials by two adjacent nuclei. In the limiting case, if a chain of N atoms with N valence electrons is so long that we can assume that we have a system of N independent electrons (with N independent deep wells), then the energy levels are degenerate N times. In this case they are indiscernible from one another as they are all the same, and are denoted E_{loc} in the figure.

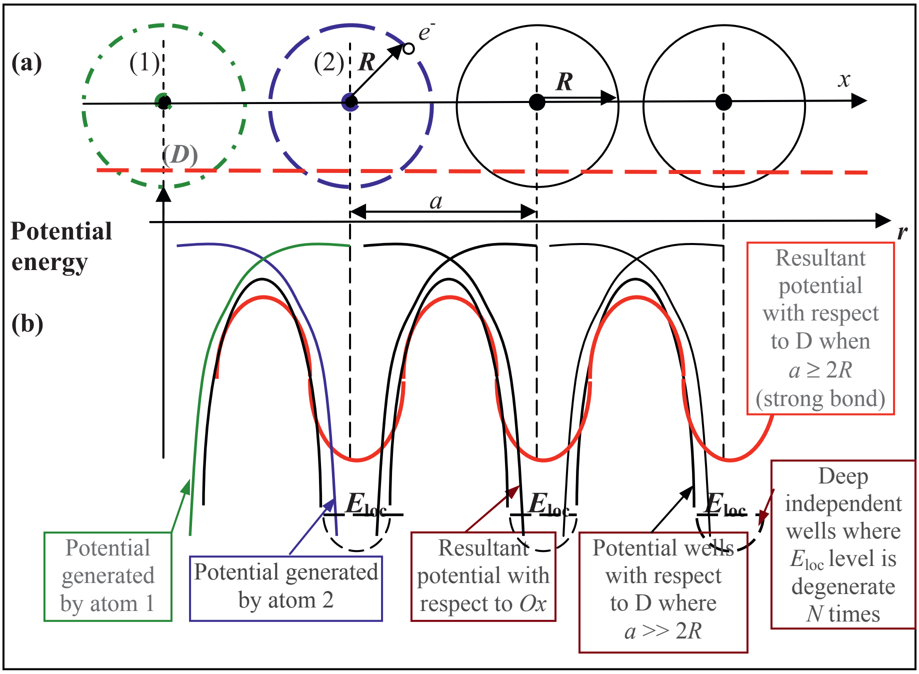


Figure 1.2. *Strong bonds and: (a) atomic orbitals (s orbitals with radius R) in a periodic lattice (of period denoted a) where $a \geq 2R$; (b) in 1D, the resulting potential energy (thick curve) seen by electrons*

– If $a \geq 2R$, the closeness of neighboring atoms induces a slight overlap of nuclei generated potentials. This means that the potential wells are no longer independent and their degeneration is increased. Electrons from one bond can interact with those of another bond, giving rise to a spread in the band energy levels. It is worth noting that the resulting potential wells are nevertheless considerably deeper than those in weak bonds (where $a < 2R$), so that the electrons remain more localized around their base atom. Given these well depths, the perturbation method that was used for weak bonds is no longer viable. Instead, in order to treat this system we will have to turn to the Hückel method or use Floquet's theorem (see Chapter 7).

1.3.4. *Choosing between approximations for weak and strong bonds*

The electrical behavior of metals is essentially determined by that of the conduction electrons. As detailed in section 1.3.2, these electrons are delocalized throughout the whole lattice and should be treated as weak bonds.

Dielectrics (insulators), however, have electrons which are highly localized around one or two atoms. These materials can therefore only be described using strong-bond theory.

Semiconductors have carriers which are less localized. The external electrons can delocalize over the whole lattice, and can be thought of as semi-free. Thus, it can be more appropriate to use the strong-bond approximation for valence electrons from the internal layers, and the weak-bond approximation for conduction electrons.

1.4. Complementary material: basic evidence for the appearance of bands in solids

This section will be of use to those who have a basic understanding of wave mechanics or more notably experience in dealing with potential wells. For others, it is recommended that they read the complementary sections at the end of Chapters 2 and 3 beforehand.

This section shows how the bringing together of two atoms results in a splitting of the atoms' energy levels. First, we associate each atom with a straight-walled potential well in which the electrons of each atom are localized. Second, we recall the solutions for the straight-walled potential wells, and then analyze their change as the atoms move closer to one another. It is then possible to imagine without difficulty the effect of moving N potential wells, together representing N atoms making up a solid.

1.4.1. *Basic solutions for narrow potential wells*

In Figure 1.3, we have $W > 0$, and this gives potential wells at intervals such that $[-a/2, +a/2]$ where $-W < 0$.

We can thus state that $W = \frac{\hbar^2\gamma^2}{2m} > 0$, and the energy E is the sum of kinetic energy ($E_c = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m}$) and potential energy.

As the related states are carry electrons then $E < 0$, and we can therefore write that:

$$E = -W + E_c = -\frac{\hbar^2}{2m}\left(\gamma^2 - k^2\right) = -\frac{\hbar^2\alpha^2}{2m} < 0.$$

By making $\alpha^2 = \left(\gamma^2 - k^2\right) > 0$, α is real.

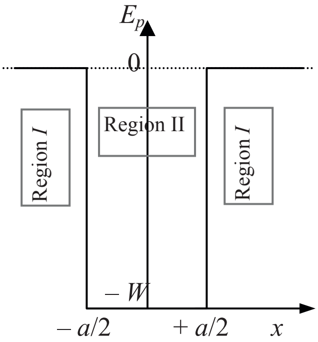


Figure 1.3. *Straight potential wells of width a*

Schrödinger's equation $\frac{d^2\psi}{dx^2}+\frac{2m}{\hbar^2}(E-V)\psi=0$ (where V is the potential energy such that $V = -W$ between $-a/2$ and $a/2$ but $V = 0$ outside of the well) can be written for the two regions:

$$\text{– region I for } |x| > \frac{a}{2} : \frac{d^2\psi}{dx^2}+\frac{2m}{\hbar^2}E\psi=0, \qquad [1.8]$$
$$\text{so that } \frac{d^2\psi}{dx^2}-\alpha^2\psi=0$$

$$\text{– region II for } -\frac{a}{2} \le x \le \frac{a}{2} : \frac{d^2\psi}{dx^2}+\frac{2m}{\hbar^2}(E+W)\psi=0, \qquad [1.9]$$
$$\text{so that } \frac{d^2\psi}{dx^2}+k^2\psi=0.$$

The solutions for equation [1.8] are (with the limiting conditions of $\psi(x)$ being finite when $x \to \pm\infty$):

– for $x > \frac{a}{2} : \psi_I(x) = Ae^{-\alpha x}$

– for $x < -\frac{a}{2} : \psi_I(x) = Ae^{+\alpha x}$.

The solution to equation [1.9] must be stationary because the potential wells are narrow (which forbids propagation solutions). There are two types of solution:

– a symmetric solution in the form $\psi_{\text{II}}(x) = B\cos kx$, for which the conditions of continuity with the solutions of region I give:

$$\left.\begin{aligned}\psi_I\left(\frac{a}{2}\right) &= \psi_{II}\left(\frac{a}{2}\right)\\ \left[\frac{\partial\psi_I}{\partial x}\right]_{\left(x=\frac{a}{2}\right)} &= \left[\frac{\partial\psi_{II}}{\partial x}\right]_{\left(x=\frac{a}{2}\right)}\end{aligned}\right\} \quad [1.10]$$

from which it can be deduced that $ka \tan\frac{1}{2}ka = \alpha a$,

– an asymmetric solution in the form $\psi_{\text{II}}(x) = B\sin kx$. Just as before, the conditions of continuity make it possible to obtain the relationship written:

$$ka \ \text{cotan}\frac{1}{2}ka = -\alpha a. \quad [1.11]$$

Equations [1.10] and [1.11] can be combined in the form:

$$\tan ka = \frac{2\alpha k}{k^2 - \alpha^2}. \quad [1.12]$$

In addition, equations [1.10] and [1.11] must be compatible with the equations that define α and *k*, so that:

$$\alpha^2 + k^2 = \gamma^2 = \frac{2mW}{\hbar^2}. \quad [1.13]$$

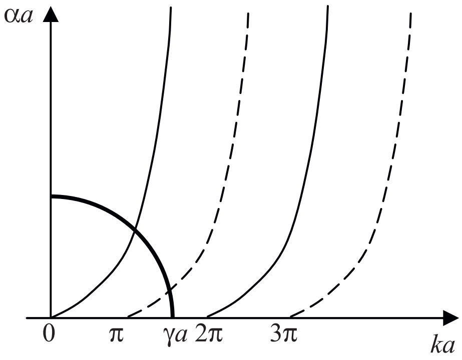


Figure 1.4. *Solutions for narrow potential wells*

The problem is normally resolved graphically. This involves noting the points where $\alpha a = f(ka)$ at the intersection of the curves described by equations [1.10] and [1.11] with the curve given by equation [1.13] (the quarter circle). The latter equation can be rewritten as:

$$\left[(\alpha a)^2 + (ka)^2\right]^{1/2} = (\gamma a). \qquad [1.14]$$

The solutions for α and thus E (as $E = -\frac{\hbar^2\alpha^2}{2m}$) correspond to the points where the circle of equation [1.14] crosses with the deduced curves from equations [1.10] and [1.11].

If:

– $\gamma a < \pi \rightarrow$ one symmetric solution $\rightarrow$ one energy level

$$– \left.\begin{array}{r}\pi < \gamma a < 2\pi \rightarrow \text{two asymmetric solution}\\ +\ \text{two asymmetric solution}\end{array}\right\} \rightarrow \text{two energy levels}$$

– $(n-1)\pi < \gamma a < n\pi \rightarrow n$ energy levels

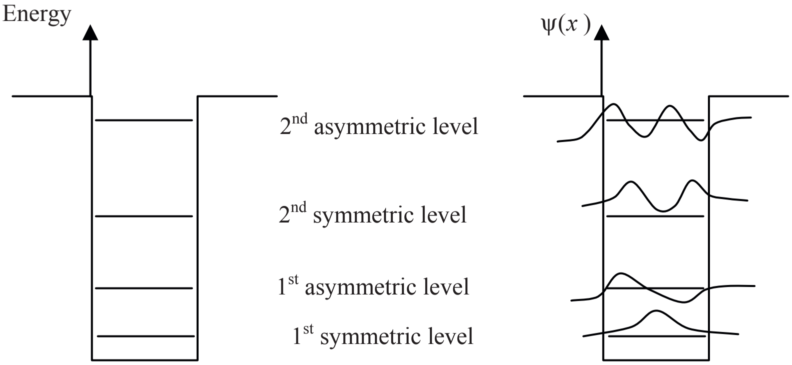


Figure 1.5. *The first four energy levels along with the appearance of the corresponding wave functions in the narrow potential wells*

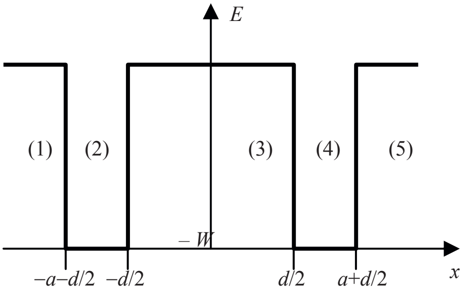


Figure 1.6. *Scheme of the potential energies of two narrow potential wells brought close to one another*

1.4.2. *Solutions for two neighboring narrow potential wells*

Schrödinger's equation, written for each of the regions denoted 1 to 5 in Figure 1.6 gives the following solutions (which can also be found in the problems later on in the book):

– Symmetric solution:

$$\psi_1 = Ce^{\alpha x}$$
$$\psi_2 = B \cos(kx + \varphi)$$
$$\psi_3 = A \, ch\alpha x$$
$$\psi_4 = B \cos(kx - \varphi)$$
$$\psi_5 = Ce^{-\alpha x}.$$

Conditions of continuity for $x = d/2$ and for $x = d/2 + a/2$ make it possible to state that:

$$\frac{\alpha}{k} = \frac{\tan ka - \frac{\alpha}{k} \, th\alpha\frac{d}{2}}{1 + \frac{\alpha}{k}\tan ka \; th\alpha\frac{d}{2}}; \qquad [1.15]$$

– Similarly, for the asymmetric solution we obtain:

$$\frac{\alpha}{k} = \frac{\tan ka - \frac{\alpha}{k} \, \coth\alpha\frac{d}{2}}{1 + \frac{\alpha}{\mathrm{k}}\tan ka \; \coth\alpha\frac{d}{2}}. \qquad [1.16]$$

1.4.2.1. *Neighboring potential wells that are well separated*

If *d* is very large, equations [1.15] and [1.16] become:

$$\frac{\alpha}{k} = \frac{\tan ka - \frac{\alpha}{k}}{1+\frac{\alpha}{\mathrm{k}}\tan ka} \qquad [1.17]$$

and tend to give the same solutions as those obtained above for narrow wells. In effect, by making $\frac{\alpha}{k} = \tan\theta$, equation [1.17] is then written as $\tan\theta = \tan\left(ka - \theta\right)$ for which the solution is $\theta = \frac{1}{2}\left(ka + n\pi\right)$. This in turn gives:

– if *n* is even then $\frac{\alpha}{k} = \tan\frac{ka}{2}$;

– if *n* is odd then $\frac{\alpha}{k} = -\text{cotan}\frac{ka}{2}$.

In effect, we again find the solutions of equations [1.10] and [1.11] for isolated wells, which is quite normal because when *d* is large the wells are isolated. Here though with a high value of *d*, the solution is degenerate as there are in effect two identical solutions, i.e. those of the isolated wells.

1.4.2.2. *Closely placed neighboring wells*

If *d* is small, we have $e^{-\alpha d} << 1$ and:

$$\left.\begin{aligned} \tanh\alpha\frac{d}{2} &\simeq 1 - 2e^{-\alpha d} \\ \coth\alpha\frac{d}{2} &\simeq 1 + 2e^{-\alpha d} \end{aligned}\right\} \Rightarrow \text{equations [1.15] and [1.16] give:}$$

$$\frac{\alpha}{k} = \frac{\tan ka - \frac{\alpha}{k}\left[1 - 2\mathrm{e}^{-\alpha d}\right]}{1+\frac{\alpha}{\mathrm{k}}\left(1 - 2\mathrm{e}^{-\alpha \mathrm{d}}\right)\tan\beta\mathrm{a}} \qquad [1.18]$$

$$\text{and } \frac{\alpha}{k} = \frac{tg\beta a - \frac{\alpha}{k}\left[1 + 2\mathrm{e}^{-\alpha d}\right]}{1+\frac{\alpha}{\mathrm{k}}\left(1 + 2\mathrm{e}^{-\alpha \mathrm{d}}\right)\tan\beta\mathrm{a}} \qquad [1.19]$$

For the single solution (α_0) in equation [1.17] (if the wells are infinitely separated) there are now two solutions: one is α_s from equation [1.18] and the other is α_a from equation [1.19]. For isolated or well separated potential wells, all states (symmetric or asymmetric) are duplicated with two neighboring energy states (as α_s and α_a are in fact slightly different from α_0). The difference in energy between the symmetric and asymmetric states tends towards zero as the two wells are separated $(d \rightarrow \infty)$. In addition, we can show quite clearly that the symmetric state is lower than the asymmetric state as in Figure 1.7.

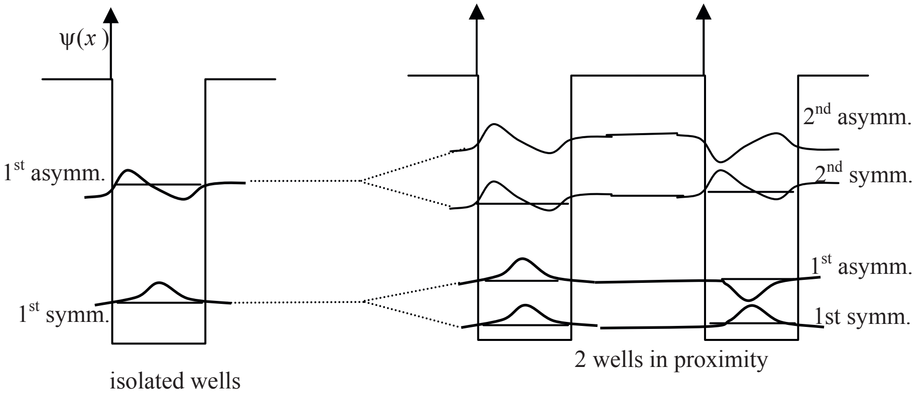


Figure 1.7. *Evolution of energy levels and electronic states on going from one isolated well to two close wells*

The example given shows how bringing together the discrete levels of the isolated atoms results in the creation of energy bands. The levels permitted in these bands are such that:

– two wells induce the formation of a "band" of two levels;

– *n* wells induce the formation of a "band" of *n* levels.

Chapter 2

The Free Electron and State Density Functions

2.1. Overview of the free electron

2.1.1. *The model*

As detailed in Chapter 1, the potential (V) (rigorously termed the potential energy) for a free electron (within the zero order approximation for solid-state electronics) is that of a flat-bottomed basin, as shown by the horizontal line in the 1D model of Figure 2.1. It can also be described by $V = V^0 = 0$.

For a free electron placed in a potential $V^0 = 0$ with an electronic state described by its proper function with energy and amplitude denoted by E^0 and ψ^0, respectively, the Schrödinger equation for amplitude is:

$$\Delta\psi^0 + \frac{2m}{\hbar^2} E^0 \psi^0 = 0. \qquad [2.1]$$

2.1.2. *Parameters to be determined: state density functions in k or energy spaces*

With:

$$k^2 = \frac{2mE^0}{\hbar^2}, \qquad [2.2]$$

equation [2.1] can be written as:

$$\Delta\psi^0 + k^2\psi^0 = 0. \quad [2.3]$$

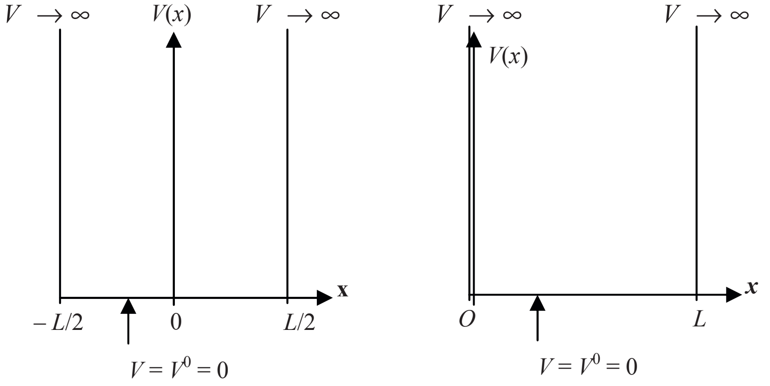


Figure 2.1. *(a) Symmetric wells of infinite depth (with the origin taken at the center of the wells); and (b) asymmetric wells with the origin taken at the well's extremity (when $0 < x < L$, we have $V(x) = 0$ so that $V(-x) = \infty$ for a 1D model)*

We shall now determine for different depth potential wells, with both symmetric and asymmetric forms, the corresponding solutions for the wave function (Ψ^0) and the energy (E^0). To each wave function there is a corresponding electronic state (characterized by quantum numbers). It is important in physical electronics to understand the way in which these states determine how energy levels are filled.

In solid-state physics, the state density function (also called the density of states) is particularly important. It can be calculated in the wave number (k) space or in the energies (E) space. In both cases, it is a function that is directly related to a dimension of space, so that it can be evaluated with respect to $L = 1$ (or $V = L^3 = 1$ for a 3D system). In k space, the state density function [$n(k)$] is such that in one dimension $n(k)\,dk$ represents the number of states placed in the interval dk (i.e. between k and $k + dk$ in k wave number space). In 3D, $n(k)\,d^3k$ represents the number of states placed within the elementary volume d^3k in k space.

Similarly, in terms of energy space, the state density function [$Z(E)$] is such that $Z(E)\,dE$ represents the number of states that can be placed in the interval dE (i.e. inclusively between E and $E + dE$ in energy space). The upshot is that if $F(E)$ is the occupation probability of a level denoted E, then the number $N(E)\,dE$ of electrons distributed in the energy space between E and $E + dE$ is equal to $N(E)\,dE = Z(E)\,F(E)\,dE$.

2.2. Study of the stationary regime of small scale (enabling the establishment of nodes at extremities) symmetric wells (1D model)

2.2.1. *Preliminary remarks*

For a symmetric well, as shown in Figure 2.1a, the Hamiltonian is such that $H(x) = H(-x)$, because $V(x) = V(-x)$ and $\frac{d^2}{dx^2} = \frac{d^2}{d\left[(-x)^2\right]}$. If I denotes the inversion operator, which changes x to $-x$, then

$$IH(x) = H(-x) = H(x).$$

$H(x)$ being invariant with respect to I, the proper functions of I are also the proper functions of H (see Chapter 1). The form of the proper functions of I must be such that $I\psi(x) = t\psi(x)$. We can thus write: $I\psi(x) = t\psi(x) = \psi(-x)$, and on multiplying the left-hand side by I, we now have:

$$\left.\begin{aligned} I\left[I\psi(x)\right] &= tI\psi(x) = t^2\psi(x) \\ &= I\left[\psi(-x)\right] = \psi(x) \end{aligned}\right\} \Rightarrow t^2 = 1, \text{ and } t = \pm 1.$$

The result is that

$$\left.\begin{aligned} I\psi(x) &= t\psi(x) = \pm\psi(x) \\ &= \psi(-x) \end{aligned}\right\} \Rightarrow \psi(x) = \pm\psi(-x). \qquad [2.4]$$

In these cases, the form of the solutions are either symmetric, as in $\psi(x) = \psi(-x)$, or asymmetric, as in $\psi(x) = -\psi(-x)$.

2.2.2. *Form of stationary wave functions for thin symmetric wells with width (L) equal to several inter-atomic distances ($L \approx a$), associated with fixed boundary conditions (FBC)*

$$\psi\left(-\frac{L}{2}\right) = \psi\left(\frac{L}{2}\right) = 0. \qquad [2.5]$$

This limiting condition is equivalent to the physical status of an electron that cannot leave the potential well due to it being infinitely high. The result is that between $x = -\frac{L}{2}$ and $x = +\frac{L}{2}$ there is a zero probability of presence, hence the preceding FBC:

$$\psi\left(-\frac{L}{2}\right) = \psi\left(\frac{L}{2}\right) = 0.$$

The general stationary solution to equation [2.3] is:

$$\psi^0(x) = A\cos kx + B\sin kx.$$

The use of the boundary conditions of equation [2.5] means that:

$$\psi^0\left(\frac{L}{2}\right) = A\cos k\frac{L}{2} + B\sin k\frac{L}{2} = 0$$

or

$$\psi^0\left(-\frac{L}{2}\right) = A\cos k\frac{L}{2} - B\sin k\frac{L}{2} = 0.$$

These last two equations result in the two same conditions:

– either $A = 0$ and $B\sin k\frac{L}{2} = 0$, so that both $\psi^0 = B\sin kx$ and $\frac{kL}{2} = n\pi$ (n is whole), so that $k = 2n\frac{\pi}{L} = N\frac{\pi}{L}$ where N is an even integer. The solution for solution ψ^0 is thus $\psi_N^0 = B\sin\frac{N\pi}{L}x$, with N being even; [2.6]

– or $B = 0$ and $A\cos k\frac{L}{2} = 0$, so that $\psi^0 = A\cos kx$ and $\frac{kL}{2} = \frac{\pi}{2} + n\pi$ (n is an integer), so that $k = \frac{\pi}{L}[2n+1] = N\frac{\pi}{L}$ where N is an odd integer. The solution for ψ^0 is thus: $\psi_N^0 = A\cos\frac{N\pi}{L}x$, with N being odd. [2.7]

The normalization condition $\int_{-L/2}^{+L/2} \left|\psi_N^0(x)\right|^2 dx = 1$ gives $A = B = \sqrt{\frac{2}{L}}$, and the two solutions in equations [2.6] and [2.7] can be brought together in:

$$\psi_N^0 = \sqrt{\frac{2}{L}} \sin\frac{N\pi}{L}\left(x + \frac{L}{2}\right), \text{ where } N = 1, 2, 3, 4, \text{ etc.} \qquad [2.8]$$

For both symmetric and asymmetric solutions, k is of the form

$$k = k_N = N\frac{\pi}{L}, \qquad [2.9]$$

where N is an odd integer and the symmetric solution and is an even integer for the asymmetric solution. Thus, N takes on successive whole values i.e. 1, 2, 3, 4, etc. The value $N = 0$ is excluded as the corresponding function $\psi_0 = B \sin k_0 x = 0$ has no physical significance (zero probability of presence). The integer values $N' = -1$, -2, -3 $(= -N)$ yield the same physical result, for the same probabilities as $\left|\psi_{N'}^0\right|^2 = \left|\psi_{-N}^0\right|^2 = \left|\psi_N^0\right|^2$. Summing up, we can say that the only values worth retaining are $N = 1, 2, 3, 4$, etc.

This quantification is restricted to the quantum number N without involving spin. As we already know, spin makes it possible to differentiate between two electrons with the same quantum number N. This is due to a projection of kinetic moment on the z axis which brings into play a new quantum number, namely $m_s = \pm\frac{1}{2}$.

We thus find that each N state can be filled by two electrons, one with a spin $m_s = +1/2$ and wave function $\left|\psi_N^0 +\right\rangle$, and the other with a spin $m_s = -1/2$ and wave function $\left|\psi_N^0 -\right\rangle$.

2.2.3. *Study of energy*

From equation [2.2] we deduce that: $E^0 = \frac{\hbar^2 k^2}{2m}$. With k given by equation [2.9], we find that the energy is quantified and takes on values given by:

$$E_N^0 = \frac{\hbar^2 k_N^2}{2m} = \frac{\hbar^2}{2m}\frac{\pi^2}{L^2}N^2 = \frac{h^2}{8mL^2}N^2, \qquad [2.10]$$

where $N = 1, 2, 3, 4$, etc.

The graphical representation of $E^0 = f(k)$ is given in Figure 2.2. The energy states (E_N^0) are associated with the electronic states denoted by wave functions in the form $\left|\psi_N^0 +\right\rangle$ and $\left|\psi_N^0 -\right\rangle$ that correspond to the spin states characterized by the quantum numbers $m_s = +\frac{1}{2}$ and $m_s = -\frac{1}{2}$, respectively.

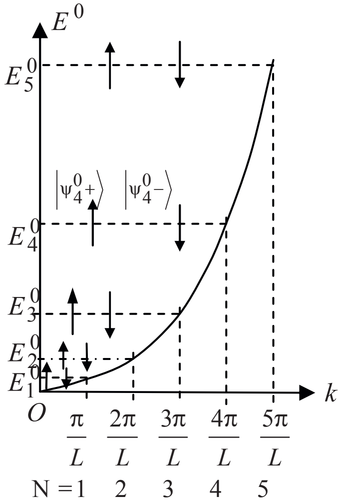


Figure 2.2. *Curve $E^0 = f(k)$ for stationary functions*

The filling of energy levels is carried out from the bottom up. The fundamental level (E_1) is filled with electrons in the states $\left|\psi_1^0+\right\rangle$ and $\left|\psi_1^0-\right\rangle$. Each level is thus filled with two electronic states that are differentiated by their spin.

2.2.4. *State density function (or "density of states") in k space*

As defined in section 2.1.2, the density of states is a function $[n(k)]$ in terms of a unit of direct space ($L = 1$) such that $n(k)\,dk$ represents the number of states that can be held within the interval dk that is between k and $k + dk$. The electronic states are represented (and numbered) by the wave functions $\left|\psi_N^0 \pm\right\rangle$.

On average, each interval of size $dk = \frac{\pi}{L}$ can hold 1 orbital (without taking spin into account). In fact, this interval corresponds to 2 orbital states; however, each is shared with the adjacent intervals. For example, in the interval $\left[\frac{2\pi}{L}, \frac{3\pi}{L}\right]$ we can place the orbital states $\left|\psi_2^0\right\rangle$ and $\left|\psi_3^0\right\rangle$, but $\left|\psi_2^0\right\rangle$ is shared with $\left[\frac{2\pi}{L}, \frac{3\pi}{L}\right]$ and $\left[\frac{\pi}{L}, \frac{2\pi}{L}\right]$, just as $\left|\psi_3^0\right\rangle$ is shared with $\left[\frac{2\pi}{L}, \frac{3\pi}{L}\right]$ and $\left[\frac{3\pi}{L}, \frac{4\pi}{L}\right]$. From this we can see that the resulting average is 1 orbital state per interval $dk = \frac{\pi}{L}$.

By taking spin into account, we can now place twice the number of states, so that in the interval $dk = \frac{\pi}{L}$ we now have two states. It is thus possible to write $n(k) \cdot \frac{\pi}{L} = 2$, so that with $L = 1$ (density function) we have:

$$n(k) = \frac{2}{\pi}. \qquad [2.11]$$

NOTE.– When $dk = 1$ we have $n(k)\, dk = n(k)$, so that the state density function $n(k)$ represents the number of states in a unit k space with the whole having a single unit dimension in direct space ($L = 1$).

2.3. Study of the stationary regime for asymmetric wells (1D model) with *L* ≈ *a* favoring the establishment of a stationary regime with nodes at extremities

The general solution for equation [2.3] is still $\psi^0(x) = A\cos kx + B\sin kx$, but the boundary conditions are now such that:

$$\psi^0(0) = \psi^0(L) = 0. \qquad [2.12]$$

From the condition $\psi^0(0) = 0$, we deduce that $A = 0$, from which $\psi^0(x) = B\sin kx$.

The condition $\psi^0(L) = 0$ thus gives $B\sin kL = 0$ where $kL = N\pi$, so that:

$$k = k_N = N\frac{\pi}{L}, \qquad [2.13]$$

in which $N = 1, 2, 3, 4$, etc. noting that $N < 0$ does not change the probability of presence; in other terms it has no physical significance. We finally arrive at:

$$\psi^0(x) = B\sin kx = B\sin k_N x = B\sin\frac{N\pi}{L}x = \psi_N^0(x).$$

B can be determined using the normalization condition, as in: $\int_0^L \left|\psi_N^0(x)\right|^2 dx = 1$ that which gives $B = \sqrt{\frac{2}{L}}$, from which

$$\psi_N^0(x) = \sqrt{\frac{2}{L}}\sin\frac{N\pi}{L}x. \qquad [2.14]$$

This solution (with $N = 1, 2, 3, 4$, etc.) replaces the solution for equation [2.8] for a symmetric system.

For its part, energy is still deduced from equation [2.2]. With the condition imposed by equation [2.13] on k_N, we are brought to the same expression as equation [2.10]:

$$E^0 = E_N^0 = \frac{\hbar^2}{2m}k_N^2 = \frac{\hbar^2}{2m}\frac{N^2\pi^2}{L^2} = E_1^0 N^2. \qquad [2.15]$$

The representation of $E^0 = f(k)$ is also still given by Figure 2.4 and the state density function $[n(k)]$ is the same as before, i.e. as in equation [2.11].

2.4. Solutions that favor propagation: wide potential wells where $L \approx 1$ mm, i.e. several orders greater than inter-atomic distances

2.4.1. *Wave function*

This problem can be seen as that of a wire, or rather molecular wire, with a given length (L) tied around on itself as shown in Figure 2.3.

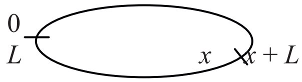


Figure 2.3. *Molecular wire of length L*

For a point with coordinate x, the probability is the same whatever the number of turns made, so we can write $\psi(x) = \psi(x+L)$. Generally, after making n turns of length L we would end up at the same point, so we can write $\psi(x) = \psi(x+nL)$ where n is an integer. The boundary condition:

$$\psi(x) = \psi(x+L) \tag{2.16}$$

is called the *periodic boundary condition* (PBC) or the Born-von Karman condition. When $x = 0$, it can be simplified so that:

$$\psi(0) = \psi(L). \tag{2.17}$$

That several revolutions are possible means that the solution must be a progressive wave. The amplitude of the free electrons wave function must take the form (see Chapter 1) given by:

$$\psi^0(r) \equiv \psi_k^0(x) = Ae^{ikx}. \tag{2.18}$$

In effect, with the form of the wave being that presented in section 1.2.2, the function $\Psi(\vec{r},t)$ here becomes the propagation wave $\Psi^0(x,t) = Ae^{i(kx-\omega t)}$, which propagates towards $x > 0$ as $k > 0$. As propagation in the opposite sense is possible, we find that $k < 0$ is therefore also physically possible.

The normalization condition $\int_0^L |\psi_k(x)|^2 dx = 1$ makes it possible to determine $A = \sqrt{\frac{1}{L}}$. For its part the condition expressed in equation [2.17] gives $e^{ik0} = e^{ikL}$, so that $1 = e^{ikL}$, from which $kL = 2N\pi$, which in turn means that:

$$k = k_N = \frac{2\pi}{L}N \text{, where } N = 0, \pm 1, \pm 2, \pm 3, \pm 4.... \tag{2.19}$$

(the solution for $N = 0$ simply gives a probability for a constant presence).

Placing these results into equation [2.18] we finally have for the amplitude function:

$$\psi_N^0(r) \equiv \psi_{k_N}^0(x) = Ae^{ik_N x} = \frac{1}{\sqrt{L}}e^{i\frac{2\pi}{L}Nx}. \tag{2.20}$$

NOTE.– We can immediately say that for the conditions that favor propagation, we now have $\left|\psi^0_{k_N}(x)\right|^2 = \psi^0_{k_N}\overline{\psi^0_{k_N}} = A^2 = \frac{1}{L}$, a constant value wherever along (x) an electron might be. The electrons move freely, without any specific localization (i.e. the probability of their presence is constant, whatever the value of x).

2.4.2. *Study of energy*

By taking the expression for k given in equation [2.19] and placing it into equation [2.2], we obtain:

$$E^0 = E^0_N = \frac{\hbar^2}{2m}\left(\frac{2\pi}{L}\right)^2 N^2, \quad \text{with } N = \pm 1, \pm 2, \pm 3, \pm 4.... \qquad [2.21]$$

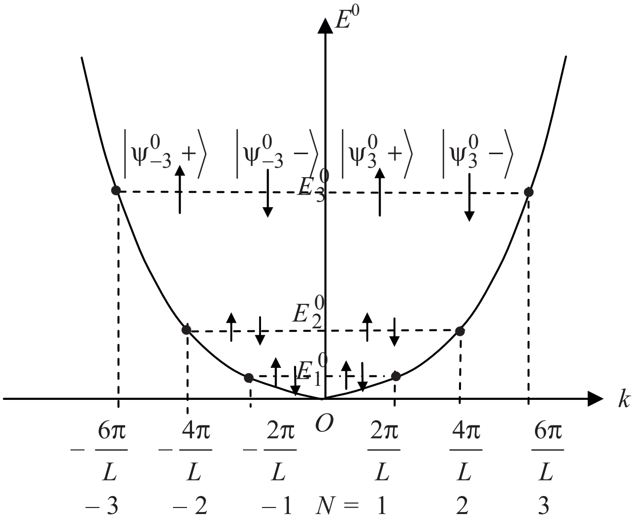


Figure 2.4. *Curve of $E^0 = f(k)$ for progressive solutions*

$E^0_N = f(k)$ is represented in Figure 2.4. Without taking spin into account, at each electronic level there are two electronic states (associated with the two possible values of N, as in $N = \pm|N|$). Including spin, each level actually corresponds to four states. To put it another way, we can state that each point on the curve corresponding to either a negative or a positive value of N is associated with two states (up and down spin).

When taking spin into account, we also find that the degree of degeneracy is four as each energy level can accommodate four electrons, each corresponding to a specific wave function. Thus, at the N^{th} level these four functions are:

$$\left|\psi^0_{|N|}\ +\right\rangle,\ \left|\psi^0_{|N|}\ -\right\rangle,\ \left|\psi^0_{-|N|}\ +\right\rangle,\ \left|\psi^0_{-|N|}\ -\right\rangle.$$

2.4.3. *Study of the state density function in k space*

Taking electron spin into account, we can now place on average two electrons into the interval $dk = \frac{2\pi}{L}$. There are four electrons in all, but with two in each division. Thus, $n(k)\cdot\frac{2\pi}{L} = 2$, so that with $L = 1$:

$$n(k) = \frac{1}{\pi}. \qquad [2.22]$$

To conclude, we can see that with progressive solutions, the number of states that can be placed in a unit interval in reciprocal space is equal to 1/π. One half of that can be placed in stationary solutions, even though the available k space is twice as large. It should be noted that the negative values of N and thus of k must also be taken into account.

2.5. State density function represented in energy space for free electrons in a 1D system

The curves given by $E = f(k)$ give a direct relation between k and energy spaces. In the space that we have defined, as detailed in section 2.1.2, the $Z(E)$ state density function is such that $Z(E)\ dE$ represents the number of energy states between E and $E + dE$ with respect to a material unit dimension (in 1D, $L = 1$ unit length).

Rigorously speaking, $Z(E)$ should be a discontinuous function as it is defined, *a priori*, only for discrete values of energy corresponding to the solution of the

Schrödinger equations [2.10] and [2.15] or [2.21] as below, respectively for stationary or progressive cases:

$$E_N^0 = \frac{\hbar^2 k_N^2}{2m} = \frac{\hbar^2}{2m}\frac{4\pi^2}{L^2}N^2 = \frac{h^2}{2mL^2}N^2. \qquad [2.21]$$

A numerical estimation can be carried out to find the typical value for free (conduction) electrons and, in this example, shows that $E_F \approx 3$ eV (Fermi energy measured with respect to the bottom of the potential wells).

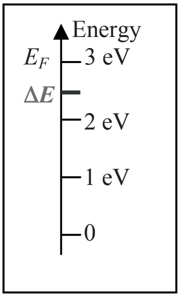
By making $L = 1$ mm (which is small enough to be at the scale typically used for lab samples, and large enough to contain a sufficiently high enough number of particles to give a statistical average), equation [2.18] gives $N \approx 1.5 \times 10^6$.

The difference in energy between two adjacent states $[N+1]$ and N is thus given by:

$$\Delta E = E_{N+1} - E_N = \frac{h^2}{2mL^2}(2N+1) \approx \frac{h^2}{mL^2}N,$$

giving $\Delta E \approx 4 \times 10^{-6}$ eV.

This holds where E_F is small and in effect the conduction electrons show a discrete energy value that can be neglected in an overall representation of electron energies (see below).



To conclude, the energy levels are quantified but the difference in their energies are so small that the function $Z(E)$ as defined above can be considered as being quasi-continuous around E_F. Often the term quasi-continuum is used in this situation.

2.5.1. *Stationary solution for FBC*

Here, as shown in Figure 2.2, only values with $k > 0$ are physically relevant. E_1^0 pertains to a single value (k_1) in k space. Similarly, E_2^0 corresponds to a single value (k_2). A consequence of this relationship between energy and k spaces is that for a number of states with energies between E and $E + dE$ there is a corresponding and equal number of states between k and $k + dk$. This can be written as:

$$Z(E)\, dE = n(k)\, dk. \qquad [2.23]$$

From this it can be deduced that $Z(E) = \dfrac{n(k)}{\frac{dE}{dk}}$.

With E from equation [2.2], i.e. $E = \dfrac{\hbar^2}{2m} k^2$, we also equally have $k = \dfrac{\sqrt{2mE}}{\hbar}$ and thus $\dfrac{dE}{dk} = \dfrac{\hbar^2}{2m} 2k = \dfrac{\hbar^2}{m}\dfrac{\sqrt{2mE}}{\hbar} = \hbar\left(\dfrac{2}{m}\right)^{1/2} E^{1/2}$.

From this it can be deduced that, for $n(k)$ given by equation [2.11] (or rather $n(k) = \frac{2}{\pi}$):

$$Z(E) = \frac{1}{\pi\hbar}\sqrt{\frac{2m}{E}}. \qquad [2.24]$$

Thus, when E increases, $Z(E)$ decreases, as shown in Figure 2.5.

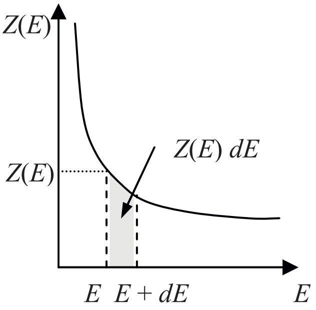


Figure 2.5. *State density function for a stationary or progressive 1D system*

2.5.2. *Progressive solutions for progressive boundary conditions (PBC)*

As shown in Figures 2.4 and 2.6, here the interval dE corresponds simultaneously to the intervals dk_+ (for $k > 0$) and dk_- (for $k < 0$).

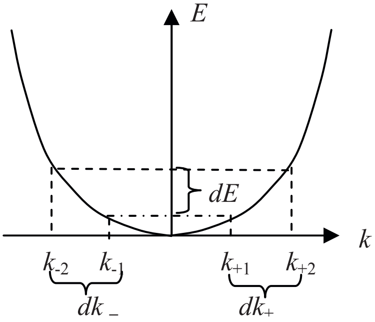


Figure 2.6. *Plot of E = f(k) for progressive solutions*

As in both dk_+ and dk_- we can place the same number of states, it is possible to state that:

$$Z(E)dE = n(k)dk_+ + n(k)dk_- = 2n(k)dk.$$

Thus, $Z(E) = 2\frac{n(k)}{\frac{dE}{dk}}$.

With $n(k)$ given by equation [2.22], $n(k) = \frac{1}{\pi}$, we obtain: $Z(E) = \frac{2}{\pi}\frac{1}{\frac{dE}{dk}}$,

where $\frac{dE}{dk}$ was calculated in the preceding section.

We again find the same expression for $Z(E)$, as shown in equation [2.24] and thus the same graphical representation as shown in Figure 2.5.

2.5.3. *Conclusion: comparing the number of calculated states for FBC and PBC*

Stationary waves: FBC

$$E_N^0 = \frac{\hbar^2}{2m}\frac{\pi^2}{L^2}N^2$$

$$n(k) = \frac{2}{\pi}$$

Progressive waves: PBC

$$E_N^0 = \frac{\hbar^2}{2m}\left(\frac{2\pi}{L}\right)^2 N^2$$

$$n(k) = \frac{1}{\pi}$$

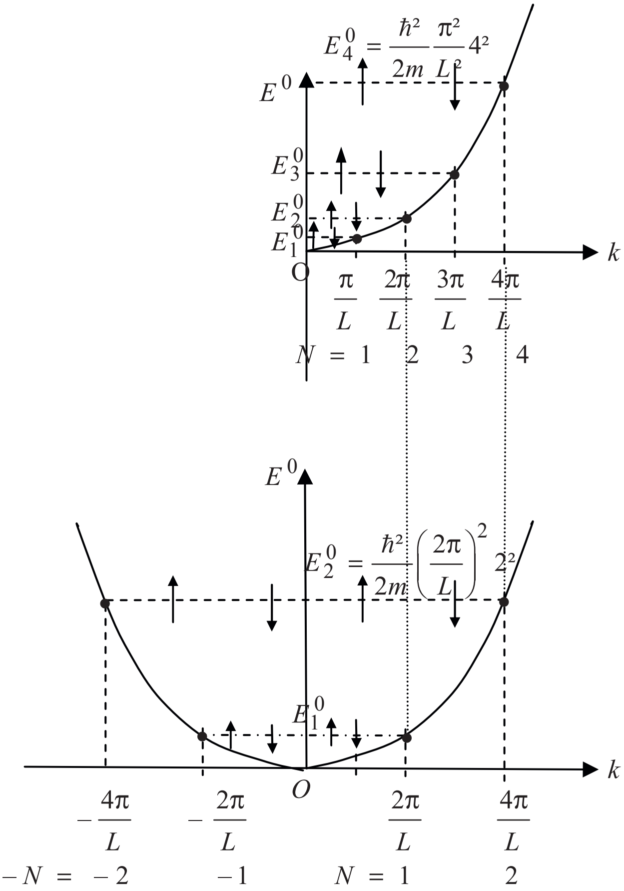


Figure 2.7. *FBC and PBC states*

The total number of states calculated over all the reciprocal space is in fact the same for the two types of solution, as with the FBC there is the involvement of only one half-space and $n(k) = 2/\pi$, while under PBC both half-spaces are brought in (i.e. $k > 0$ and $k < 0$) and $n(k) = 1/\pi$. It can be seen in Figure 2.7 that there are eight states represented when taking into account spin for the two cases (four states not accounting for spin associated with the points on the plots in Figure 2.7) with:

– k varying from 0 to $4\pi/L$ or from 0 to $\pm 4\pi/L$;

– E varying from 0 to $\left[E_4^0\right]_{FBC} = \frac{\hbar^2}{2m}\frac{\pi^2}{L^2}4^2 = \frac{\hbar^2}{2m}\left(\frac{2\pi}{L}\right)^2 2^2 = \left[E_2^0\right]_{PBC}$.

2.6. From electrons in a 3D system (potential box)

2.6.1. *Form of the wave functions*

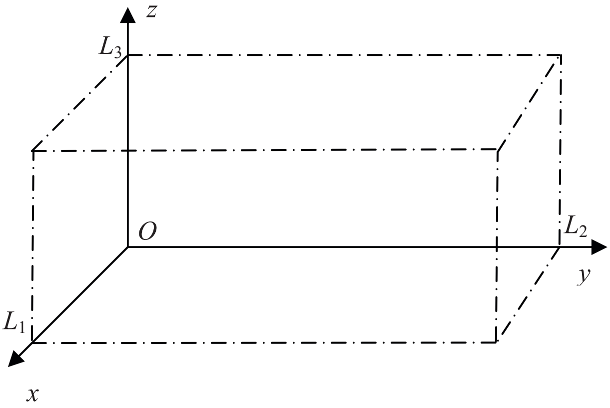


Figure 2.8. *A parallelepiped box (direct space)*

We assume that the free electrons are closed within a parallelepiped box with sides of length L_1, L_2, L_3 as shown in Figure 2.8. The potential is zero inside the box and infinite outside. The Schrödinger equation is thus given by:

$$\Delta\psi^0(x,y,z) + \frac{2m}{\hbar^2}E^0\psi^0(x,y,z) = 0.$$

By making $k^2 = \frac{2m}{\hbar^2}E^0$, it can be rewritten as: $\Delta\psi^0 + k^2\psi^0 = 0$.

The resolution of this equation can be carried out after separating the variables. In order to do this, we make: $\psi^0(x,y,z) = \psi_1^0(x)\psi_2^0(y)\psi_3^0(z)$ and $E^0 = E_1^0 + E_2^0 + E_3^0$. From this we can deduce three equations with the same form:

a) $\frac{d^2\psi_1^0(x)}{dx^2} + \frac{2m}{\hbar^2}E_1^0\psi_1^0(x) = 0$, so that on making $k_x^2 = \frac{2m}{\hbar^2}E_1^0$ we have:

$$\frac{d^2\psi_1^0(x)}{dx^2} + k_x^2\psi_1^0(x) = 0$$

b) $\frac{d^2\psi_2^0(y)}{dy^2} + \frac{2m}{\hbar^2}E_2^0\psi_2^0(y) = 0$, so that on making $k_y^2 = \frac{2m}{\hbar^2}E_2^0$ we have:

$$\frac{d^2\psi_2^0(y)}{dy^2} + k_y^2\psi_2^0(y) = 0$$

c) $\frac{d^2\psi_3^0(z)}{dz^2} + \frac{2m}{\hbar^2}E_3^0\psi_3^0(z) = 0$, so by making $k_z^2 = \frac{2m}{\hbar^2}E_3^0$ we have:

$$\frac{d^2\psi_3^0(z)}{dz^2} + k_z^2\psi_3^0(z) = 0$$

2.6.1.1. *Case favoring fixed boundary conditions*

Here the FBCs are:

– with respect to *Ox*: $\psi_1^0(0) = \psi_1^0(L_1) = 0$,

– with respect to *Oy*: $\psi_2^0(0) = \psi_2^0(L_2) = 0$,

– with respect to *Oz*: $\psi_3^0(0) = \psi_3^0(L_3) = 0$.

The use of these boundary limits means that we can solve these differential equations directly from the equivalent 1D system (the boundary limits are identical to those in the 1D system with an origin at an extremity – see section 2.3):

$$\begin{aligned}\psi^0(x,y,z) = \psi_n^0(x,y,z) &= \psi_{n_x}^0(x)\psi_{n_y}^0(y)\psi_{n_z}^0(z)\\ &= A\sin\left[\frac{n_x\pi}{L_1}x\right]\sin\left[\frac{n_y\pi}{L_2}y\right]\sin\left[\frac{n_x\pi}{L_3}z\right],\end{aligned} \qquad [2.25]$$

where $k_x = \frac{\pi}{L_1}n_x$, $k_y = \frac{\pi}{L_2}n_y$, $k_z = \frac{\pi}{L_3}n_z$ and n_x, n_y and n_z are positive integers.

Energy is given by:

$$\begin{aligned}E^0 &= E_1^0 + E_2^0 + E_3^0 = \frac{\hbar^2}{2m}\left[k_x^2 + k_y^2 + k_z^2\right]\\ &= \frac{\hbar^2}{2m}\left[\frac{\pi^2}{L_1^2}n_x^2 + \frac{\pi^2}{L_2^2}n_y^2 + \frac{\pi^2}{L_3^2}n_z^2\right].\end{aligned} \qquad [2.26]$$

2.6.1.2. *Case favoring progressive boundary conditions*

Analogously to the case for limiting conditions, we have:

– with respect to *Ox*: $\psi_1^0(x) = \psi_1^0(x + L_1)$,

– with respect to *Oy*: $\psi_2^0(y) = \psi_2^0(y + L_2)$,

– with respect to *Oz*: $\psi_3^0(z) = \psi_3^0(z + L_3)$.

The use of these boundary limits means that we can solve these differential equations directly from the equivalent 1D system (the boundary limits are identical to those in the 1D system where propagation is favored – see section 2.4.1):

$$\begin{aligned}\psi_n^0(r) \equiv \psi_k^0(x,y,z) &= \psi_{k_x}^0\psi_{k_y}^0\psi_{k_z}^0\\ &= A_{1x}A_{2y}A_{3z}e^{ik_x x}e^{ik_y y}e^{ik_z z} = Ae^{i\frac{2\pi}{L_1}n_x x}e^{i\frac{2\pi}{L_2}n_y y}e^{i\frac{2\pi}{L_3}n_z z}\end{aligned} \qquad [2.27]$$

in which $A_{1x}A_{2y}A_{3z} = A$, $k_x = \frac{2\pi}{L_1}n_x$, $k_y = \frac{2\pi}{L_2}n_y$, $k_z = \frac{2\pi}{L_3}n_z$ and where n_x, n_y and n_z are positive or negative integers.

Energy is given by:

$$E^0 = E_1^0 + E_2^0 + E_3^0 = \frac{\hbar^2}{2m}\left[k_x^2 + k_y^2 + k_z^2\right] = \frac{h^2}{2m}\left[\frac{n_x^2}{L_1^2} + \frac{n_y^2}{L_2^2} + \frac{n_z^2}{L_3^2}\right]. \quad [2.28]$$

2.6.2. *Expression for the state density functions in k space*

2.6.2.1. *Where stationary solutions are favored*

In 3D, we can divide the k space into elemental cells (n_x, n_y, n_z, which change values as integers) such that their smallest variation is given by $\Delta n_x = \Delta n_y = \Delta n_z = 1$ and that the sides of the smallest, elemental cell are given by $\Delta k_x = \frac{\pi}{L_1}$, $\Delta k_y = \frac{\pi}{L_2}$, $\Delta k_z = \frac{\pi}{L_3}$. The elemental cells are such that their nodes (upmost point) are associated with an electronic state represented by a wave function given by:

$$\psi^0(x,y,z) = \psi_n^0(x,y,z)$$
$$= A \sin\left[\frac{n_x \pi}{L_1}x\right]\sin\left[\frac{n_y \pi}{L_2}y\right]\sin\left[\frac{n_x \pi}{L_3}z\right] = \psi_{k_x}^0 \psi_{k_y}^0 \psi_{k_z}^0 .$$

Each elemental cell thus has eight nodes, each of which corresponds to eight states that are, in turn, each shared across eight elemental cells. Without taking spin into account, we can assert that $\frac{8}{8} = 1$ state per cell on average. Taking spin states into account we can now place two electronic states into each elemental cell.

The elemental cells have an elemental volume given by:

$$\Delta k^3 = \Delta k_x \ \Delta k_y \ \Delta k_z = \frac{\pi}{L_1}\frac{\pi}{L_2}\frac{\pi}{L_3} = \frac{\pi^3}{V},$$

where V is the volume in direct space (starting with a parallelepiped, see Figure 2.8). For a unit volume in direct space ($V = 1$ given for the calculation of the state density function), we have $\left[\Delta k^3\right]_{V=1} = \pi^3$.

Still using $n(k)$ to denote the state density function, we can write that $n(k)\left[\Delta k^3\right]_{V=1} = 2$, so that $n(k)\pi^3 = 2$. In other words,

$$n(k) = \frac{2}{\pi^3}. \qquad [2.29]$$

2.6.2.2. *Where progressive solutions are favored*

In 3D, we have $\Delta k_x = \frac{2\pi}{L_1}$, $\Delta k_y = \frac{2\pi}{L_2}$, $\Delta k_z = \frac{2\pi}{L_3}$, and the elemental cells are such that each has a node associated with an electronic state represented by a wave function in the form:

$$\psi_n^0(r) = \psi_{k_x}^0 \psi_{k_y}^0 \psi_{k_z}^0 = A e^{i\frac{2\pi}{L_1}n_x x} e^{i\frac{2\pi}{L_2}n_y y} e^{i\frac{2\pi}{L_3}n_z z}.$$

The elemental cells thus have an elemental volume given by:

$$\Delta k^3 = \Delta k_x \; \Delta k_y \; \Delta k_z = \frac{2\pi}{L_1}\frac{2\pi}{L_2}\frac{2\pi}{L_3} = \frac{8\pi^3}{V}$$

where V is the direct space volume. For a direct space unit volume ($V = 1$), we have $\left[\Delta k^3\right]_{V=1} = 8\pi^3$, and then by using $n(k)$ to denote the state density function, we can write that $n(k)\left[\Delta k^3\right]_{V=1} = 2$, so that $n(k)8\pi^3 = 2$, which in other words means:

$$n(k) = \frac{1}{4\pi^3}. \qquad [2.30]$$

2.6.3. *Expression for the state density functions in k space*

2.6.3.1. *Where stationary solutions are favored*

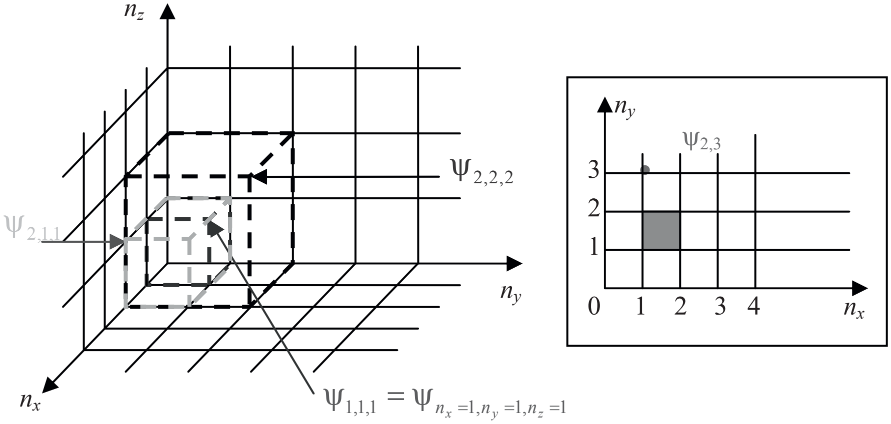
The calculation for the state density function now denoted $Z(E)$ can be carried out using two different routes:

– using the correlation between k and energy spaces;

– via a direct calculation in energy space using quantum numbers.

Here we will use the direct method.

First we can note that with stationary solutions where quantum numbers n_x, n_y and n_z are positive, the quantum number space must be held within the first octet (n_x, n_y, $n_z > 0$) as shown in Figure 2.8a. If the problem is dealt with in 2D only, then the space (or more exactly the plain) should be within the first quadrant, as indicated in Figure 2.8b.



(a) (b)

Figure 2.8. *Nodes and states in quantum number space: (a) 3D space with n_x, n_y, and n_z as whole integers; (b) likewise in 2D*

In quantum number space, just as in k space, the cells and their nodes are associated with electronic states that are characterized by their electronic wave functions, as in $\left|\psi_{n_x, n_y, n_z}\right\rangle$ where each is denoted with respect to its specific quantum number n_x, n_y, and n_z. Once we take spin into account, characterized by the quantum numbers $m_s = +\frac{1}{2}$ and $m_s = -\frac{1}{2}$, these functions in fact give rise to the electronic states $\left|\psi_{n_x,n_y,n_z}+\right\rangle$ and $\left|\psi_{n_x,n_y,n_z}-\right\rangle$.

From equation [2.23], we can see that the states of the equi-energy E = constant are spread out over a sphere of radius given by n = constant $= \sqrt{\frac{8mL^2}{h^2}E}$. Also, in the quantum number space shown in Figure 2.9, the state that have energies between E and $E + dE$ are spread between the spheres of radius n and $n + dn$ such that $n = \sqrt{\frac{8mL^2}{h^2}E}$ and $n + dn = \sqrt{\frac{8mL^2}{h^2}(E + dE)}$. This gives a volume which is equal to $4\pi n^2 dn$. With the quantum numbers being obligatorily positive we have restricted the quantum number space to the first octant (where n_x, n_y, $n_z > 0$) in which the preceding volume (divided by eight) is reduced to $\frac{4\pi n^2 dn}{8} = \frac{\pi n^2 dn}{2}$.

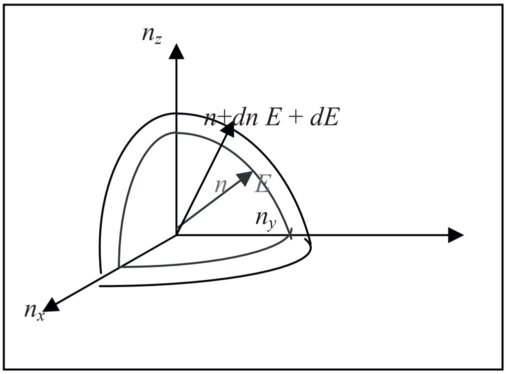


Figure 2.9. *The relation between the quantum number and energy spaces*

The number of cells that can be placed in this volume is given simply by $\frac{\pi n^2 dn}{2}$ when the cells have a volume given by $\Delta n_x \, \Delta n_y \, \Delta n_z = 1^3 = 1$ (the quantum number can vary in single steps in each particular direction, as in Figure 2.8). Without taking spin into account, each cell contains eight states, but each state is shared between eight cells, so that on average we can therefore place one state in each cell. So, again without taking spin into account, the number of electronic states that can be placed is thus $\frac{\pi n^2 dn}{2}$.

Taking spin into account means that we can now place twice the number of states, so that the total number of electronic states is equal to $\pi n^2 dn$. We can thus finally write that:

$$Z(E)dE = \pi n^2 dn.$$

With equation [2.26] (where $n^2 = n_x^2 + n_y^2 + n_z^2$ and it is supposed that $L_1 = L_2 = L_3 = 1$ so as to have a unit volume from which the density functions can be obtained) which takes the form of $E = \frac{h^2}{8mL^2} n^2 = \frac{h^2}{8m} n^2$, we have $dE = \frac{h^2}{4m} ndn$, from which $\frac{4m}{h^2} dE = ndn$.

The result in that: $\pi n^2 dn = \sqrt{\frac{8m}{h^2} E}\, \pi \frac{4m}{h^2} dE = Z(E)dE$, so that:

$$Z(E) = \frac{4\pi(2m)^{3/2}}{h^3} \sqrt{E}. \qquad [2.31]$$

The curve thus has a parabolic shape, as shown in Figure 2.10.

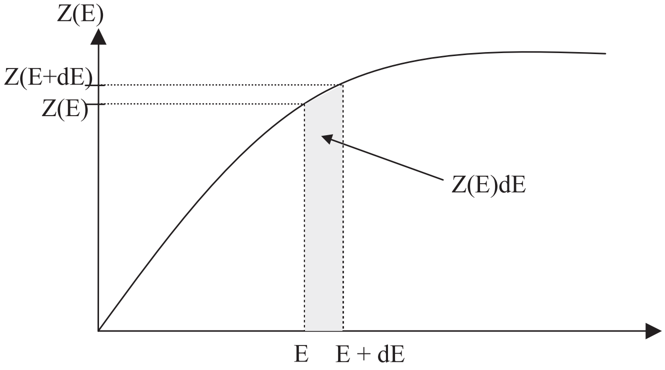


Figure 2.10. *State density function in 3D*

NOTE.– If we use the correlation between k space and energy space, we can write:

$$Z(E)dE^{3D} = n(k)\frac{dV_k}{8} = n(k)\frac{4\pi k^2 dk}{8} \overset{n(k)=\frac{2}{\pi^3}}{=} \frac{1}{\pi^2}k^2 dk.$$

With $k^2 = \frac{2mE}{\hbar}$ we have on one side $k = \frac{\sqrt{2mE}}{\hbar}$, and on the other $kdk = \frac{2m}{\hbar}dE$, from which we once again obtain equation [2.28].

2.6.3.2. *Problem: where progressive solutions are favored (see also problems 5 and 6 of this chapter)*

In this case, n_x, n_y, and n_z are positive or negative integers and the quantum number spaces – just as that for k – is no longer restricted to the first octet but covers all space.

By using the correlation between k space and energy space, we can thus write:

$$Z(E)dE^{3D} = n(k)dV_k = n(k)4\pi k^2 dk \overset{n(k)=\frac{1}{4\pi^3}}{=} \frac{1}{\pi^2}k^2 dk,$$

from which we again find equation [2.31].

2.7. Problems

The reader is advised that if he or she has not yet looked at the basics of statistical thermodynamics, including the use and significance of the Fermi-Dirac function, problems 1, 3 and 4 should be attempted after reading section 4.4.3.2 of this book.

2.7.1. *Problem 1: the function Z(E) in 1D*

Here we are looking at free electrons in a small one-dimensioned medium with L being equal to several nanometers. The total number of electrons is equal to N_t and the filling of the energy levels at absolute zero is under consideration.

1) What type of boundary conditions should we use?

2) Give the value of N of the last occupied level.

3)

a) If $E_F(0) \equiv E_F$ represents the energy of the highest fully occupied level, express the value of E_F as a function of N_t.

b) From this and as a function of E_F, deduce the expression of the total numbers of orbitals [$N(E_F)$] (of electronic states including spin) for electrons with energies lower than E_F.

4)

a) Generalize the last expression for any level E that is lower than E_F.

b) From this deduce the expression for the function [$Z(E)$] of the density of energy states.

Answers

1) The medium is of a sufficiently small size so that we can assume that there is a stationary regime and as a consequence the use of fixed boundary conditions, which takes into account the presence of a node at each extremity of the system.

2) Being at absolute zero, the levels are filled as a continuum from the lowest level $N = 1$ to the highest. With two electrons placed into each level (each electron has a different spin), the highest level can thus be discerned using $N = \frac{N_t}{2}$.

3)

a) We can now write that $E_F = \left[E_N\right]_{N = N_t/2} \overset{(10)}{=} \frac{\hbar^2}{2m}\frac{\pi^2}{L^2}\left[\frac{N_t}{2}\right]^2 = \frac{\hbar^2}{2m}\left[\frac{\pi N_t}{2L}\right]^2$.

b) We thus have $N(E_F) = N_t$, so that by taking the preceding equation into account, we now have:

$$N(E_F) = \frac{2L}{\pi}\left[\frac{\hbar^2}{2m}\right]^{-1/2} E_F^{1/2}.$$

4)

a) For any level E which is such that $E < E_F$, the preceding relationship retains its validity when the level E_F is replaced by the level E (which accords to the same filling rules as the level E_F). The total number of electronic states with energy less than E, including spin states and denoted $N(E)$ is thus given by:

$$N(E) = \frac{2L}{\pi}\left[\frac{\hbar^2}{2m}\right]^{-1/2} E^{1/2}$$

b) The function $Z(E)$ is such that $[dN(E)]_{L=1} = Z(E)\,dE$ (number of states with energies comprised between E and $E = dE$ for a crystal with a unit dimension, i.e. $L = 1$), so that:

$$Z(E) = \left[\frac{dN(E)}{dE}\right]_{L=1} = \left(\frac{L}{\pi}\right)_{L=1}\left[\frac{\hbar^2}{2m}\right]^{-1/2} E^{-1/2} = \frac{1}{\pi\hbar}(2m)^{1/2} E^{-1/2}.$$

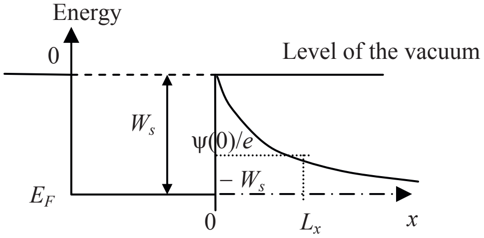
Here we return to the expression given in equation [2.24] and plotted in Figure 2.5.

2.7.2. *Problem 2: diffusion length at the metal-vacuum interface*

The electronic representation of a metal using potential bowls with flat bottoms makes it possible to define the work function (W_s) (of electrons) of the metal. The origin of the potentials was taken as being the level of electrons in a vacuum at an infinite distance from the metal with the highest energy level (last occupied level) being in the metal (at absolute zero temperature) and situated at $-W_s$. The work function is thus equal to $\Delta W = 0 - (-W_s) = W_s$ (where W_s is positive).

This level corresponds to the Fermi level E_F, which is generally defined with respect to the last occupied band (see also section 1.3 in [MOL 06]).

The potential bowl is thus represented in the following diagram, and it is noticeable that it no longer resembles the infinitely high potential wells. The depth of the wells being finite means that the electrons situated at the Fermi level can now "penetrate" the vacuum.



For a value $W_s = 3$ eV, evaluate for $x > 0$ the penetration length L_x (of diffusion) into the vacuum for electrons at the Fermi level of the metal (divided by e for the wave function following the pathway L_x).

Answer

For electrons that are likely to be in the zone $x > 0$ where $V = 0$ with an energy corresponding to that of the level $E_F = -W_s$, we have $E - V \overset{V=0}{=} E = -W_s$. The Schrödinger equation $\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}(E - V)\psi = 0$ can thus be written for these electrons as:

$$\frac{d^2\psi}{dx^2} - \frac{2m}{\hbar^2} W_s \psi = 0.$$

By making $k^2 = \frac{2m}{\hbar^2} W_s$ (with $W_s > 0$), the solution is $\psi(x) = Ae^{kx} + Be^{-kx}$. When $x > 0$, the boundary condition $\psi(\infty) = 0$ imposes $A = 0$, so that $\psi(x) = Be^{-kx}$.

The diffusion length L_x is defined by $\psi(x) \propto e^{-\frac{x}{L_x}}$, which is such that $\frac{\psi(L_x)}{\psi(0)} = \frac{1}{e}$ (divided by *e* for the wave function following the pathway L_x), and we now have: $L_x = \frac{1}{k} = \frac{\hbar}{\sqrt{2mW_s}}$.

Numerically speaking, with $m = 0.9 \times 10^{-30}$ kg, $W_s = 3$ eV $= 3 \times 1.6 \times 10^{-19}$ *J* and with $\hbar \approx 1\,Js$, we obtain $L_x \approx 10^{-10}$ *m* = 0.1 nm.

2.7.3. *Problem 3: 2D media: state density function and the behavior of the Fermi energy as a function of temperature for a metallic state*

This section considers a square 2D metal in the *Oxy* plane with N atoms distributed with spatial periodicity with respect to both *Ox* and *Oy*. The sides of the crystal are denoted L_x and L_y and are such that $L_x = L_y = L = (N + 1)\,a \approx Na$ as N is very large ($N >> 1$). The mono-valent atoms placed at the lattice nodes each liberate a free electron, which are thus present in numbers given by $N_t = N^2$, to which there is a corresponding electronic density given by $N_e = \frac{N_t}{L^2} = \frac{N^2}{L^2}$.

1) For the reciprocal space (k space) indicate:

a) The dimension of an elementary cell.

b) The form of the equi-energy curve.

c) Which surface the electrons are placed on that have energy less than or equal to a given value of E.

d) The maximum number of electrons that can be held in a unit surface.

e) The maximum number N' of electrons with energies less than or equal to a given E, where $N' = f\,[N, a, m, E]$.

2)

a) Determine the expression of the energy states density function ($Z(E) = g[m, E]$), and conclude.

b) Calculate the Fermi energy at absolute zero $EF(0)$ defined by $EF(0) = E$maximum for $T = 0$ K. What is the relation that exists between $EF(0)$ and $Z(E)$?

c) Numerical application: using $a = 0.3$ nm, calculate the value of $EF(0)$.

d) Given that the integration is $\int_0^\infty \frac{E^p\,dE}{1+e^{\gamma\left(\frac{E}{E_0}-1\right)}} \simeq E_0^{p+1}\left[\frac{1}{p+1}+\frac{\pi^2 p}{6\gamma^2}+\ldots\right]$, obtained from $N_e = \int_0^\infty Z(E) f(E)\,dE$, where $f(E)$ is the Fermi-Dirac function given by $f(E) = \frac{1}{1+\exp\left(\frac{E-E_F(T)}{kT}\right)}$ (see section 4.4.3.2 if needed), show that the Fermi energy $EF(T)$ for a 2D crystal is strictly independent of the temperature as in $EF(T) = EF$.

Answers

1)

a) We have $(\Delta k)^2 = \Delta kx \cdot \Delta ky = (\Delta k)^2 \frac{2\pi}{L_x}\frac{2\pi}{L_y} = \left(\frac{2\pi}{L}\right)^2$.

b) The energy of free electrons is given by $E = \frac{\hbar^2 k^2}{2m}$. In the k space, where $k = \frac{\sqrt{2mE}}{\hbar^2}$, the energy is constant for $k = \frac{\sqrt{2mE}}{\hbar^2} = \text{constant}$ (equation for a circle of radius $R = k = \frac{\sqrt{2mE}}{\hbar^2}$). The cells distributed around the circle contain electrons with a given equi-energy E.

c) Electrons with energies lower than a given value E are thus spread around the inside of the circle surface with a radius $k = \frac{\sqrt{2mE}}{\hbar^2} = \text{constant}$, and surface value $\pi k^2 = \pi\frac{2mE}{\hbar^2}$.

d) In a unit surface, the number of surface cells $\left(\frac{2\pi}{L}\right)^2$ that we can place are such that: $\frac{1}{\left(\frac{2\pi}{L}\right)^2} = \left(\frac{L}{2\pi}\right)^2$. The maximum umber of electrons that can thus be placed are (with two electrons per cell) $2x\left[\frac{L}{2\pi}\right]^2$.

e) Electrons with energies less than E are distributed on the inside of the circle of radius $k = \frac{\sqrt{2mE}}{\hbar^2}$ and of surface $\pi k^2 = \pi\frac{2mE}{\hbar^2}$. On this surface, we can thus distribute a maximum number of electrons equal to:

$$N' = 2x\left[\frac{L}{2\pi}\right]^2 x\left[\pi\frac{2mE}{\hbar^2}\right] = \frac{4\pi m L^2}{h^2}E = \frac{4\pi m N^2 a^2}{h^2}E.$$

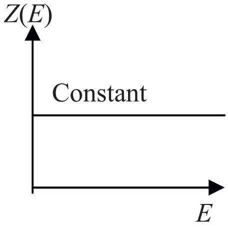
It should be noted that in order to attain the maximum number, the probability that each cell can be occupied by two electrons should be equal to unity.

2)

a) On taking spin into account, each state corresponds to an electron (non-degenerate problem). Thus, the maximum number N' of electrons with energies less than E is equal to the total number of states existing between 0 and the energy E.

The state density $Z(E)$ is such that the number per unit surface N'/L^2 of energy states below E are in accordance with: $\frac{N'}{L^2} = \int_0^E Z(E)dE$, from which $Z(E) = \frac{d(N'/L^2)}{dE}$, so that $Z(E) = \frac{4\pi m}{h^2} = \frac{m}{\pi\hbar^2} = \text{constant}$.

The function $Z(E)$ is thus a straight line in 2D space.



b) When $E = EF(0)$, all states below $EF(0)$ are occupied by $Nt = N^2$ electrons of the system, and therefore $[N']_{E=E_F(0)} = N_t = N^2$. Given the expression obtained in Question 1(e) for $E = EF(0)$, we thus have:

$$[N']_{E=E_F(0)} = \frac{4\pi m N^2 a^2}{h^2} E_F(0) = N^2,$$

from which:

$$E_F(0) = \frac{h^2}{4\pi m a^2} = \frac{\pi\hbar^2}{ma^2} = \frac{N^2}{L^2 Z(E)}.$$

c) Numerical application. With $a = 0.3$ nm, we obtain $EF(0) = 2.7$ eV.

d) With $Z(E) = \frac{N^2}{L^2 E_F(0)}$, we can thus write:

$$\frac{N^2}{L^2} = N_e = \int_0^{\infty} Z(E) f(E) dE = \frac{N^2}{L^2 E_F(0)} \int_0^{\infty} f(E) dE,$$

from which $E_F(0) = \int_0^{\infty} f(E) dE$, so that:

$$E_F(0) = \int_0^{\infty} \frac{1}{1 + e^{\frac{E_F(T)}{kT}\left(\frac{E}{E_F(T)} - 1\right)}} dE \overset{p=0, \gamma = E_F/kT}{=} E_F(T)\left[1 + 0\right] = E_F(T).$$

2.7.4. *Problem 4: Fermi energy of a 3D conductor*

Show that the Fermi energy E_F(T) of a 3D conductor described by a free electrons model is practically independent of temperature. In order to do this, consider that the total effective number of free electrons in a 3D conductor is N_e.

1) For any given temperature T, determine the relation $N_e = f(E_F(T))$ by using, just as for the 2D conductor, the fact that $N_e = \int_0^{\infty} Z(E) f(E) dE$, with $f(E) = \frac{1}{1 + \exp\left(\frac{E - E_F(T)}{kT}\right)}$ and $\int_0^{\infty} \frac{E^p dE}{1 + e^{\gamma\left(\frac{E}{E_0} - 1\right)}} \simeq E_0^{p+1}\left[\frac{1}{p+1} + \frac{\pi^2 p}{6\gamma^2} + \ldots\right]$.

2) For $T = 0$ K, find the relation $N_e = g(E_{F0})$ with $E_{F0} = (E_F)_{T=0\text{ K}}$.

3) Express E_F as a function of E_{F0}. Conclude, knowing that the value of the Fermi level for a metal is often of the order of 3 eV.

Answers

1) We have $N_e = \int_0^{\infty} Z(E) f(E) dE$, so that with $Z(E) = \frac{4\pi(2m)^{3/2}}{h^3}\sqrt{E}$ (from equation [2.28] for the 3D model) we can write $Z(E) = A_c\sqrt{E}$ with $A_c = \frac{4\pi(2m)^{3/2}}{h^3}$. Hence, we can write:

$$N_e = A_c \int_0^{\infty} \frac{E^{1/2}}{1+e^{\frac{E-E_F(T)}{kT}}} dE = A_c \int_0^{\infty} \frac{E^{1/2}}{1+e^{\frac{E_F(T)}{kT}\left(\frac{E}{E_F(T)}-1\right)}} dE$$

$$\overset{\substack{p=1/2 \\ \gamma=E_F(T)/kT}}{=} \mathrm{A_c}\left[E_F(T)\right]^{3/2}\left[\frac{2}{3}+\frac{\pi^2}{6}\frac{1}{2}\frac{k^2T^2}{E_F^2(T)}+\ldots\right].$$

2) At absolute zero, the number of electrons is unchanged. The levels are simply filled up to the Fermi level with an occupation probability of 100%. Using E_{F0} to denote the Fermi level at absolute zero, we can write that:

$$N_e = \int_0^{E_{F0}} Z(E)dE = A_c \int_0^{E_{F0}} E^{1/2} dE = \frac{2}{3} A_c E_{F0}^{3/2}.$$

3) Using the equations for N_e obtained in response to questions 1) and 2), and equating them, we obtain:

$$\left[E_F(T)\right]^{3/2}\left[\frac{2}{3}+\frac{\pi^2}{6}\frac{1}{2}\frac{k^2T^2}{E_F^2(T)}+\ldots\right] = \frac{2}{3}E_{F0}^{3/2},$$

from which

$$\left[E_F(T)\right] = E_{F0}\left[1+\frac{\pi^2}{8}\frac{k^2T^2}{E_F^2(T)}+\ldots\right]^{-2/3} \simeq E_{F0}\left[1-\frac{\pi^2}{12}\frac{k^2T^2}{E_F^2(T)}-\ldots\right]$$

$$\simeq E_{F0}\left[1-\frac{\pi^2}{12}\left(\frac{kT}{E_{F0}}\right)^2-\ldots\right].$$

In terms of numbers, when $E_{F0} \approx E_F(T) \approx 3$ eV, $T = 300$ K (so that $k = 8.6 \times 10^{-5}$ eVK^{-1} and $kT \approx 2.6 \times 10^{-2}$ eV), and the corrective term required is:

$$\frac{\pi^2}{12}\left(\frac{kT}{E_F(T)}\right)^2 \approx \frac{\pi^2}{12}\left(\frac{kT}{E_{F0}}\right)^2 \approx 5.8 \times 10^{-5} \text{ eV}.$$

The value of the corrective term is negligible with respect to the 3 eV of the metal Fermi level. We can thus assume that the position of the Fermi level is practically independent of temperature, at least within reach of ambient temperatures.

2.7.5. *Problem 5: establishing the state density function via reasoning in moment or k spaces*

Derive the aforementioned equation for the state density function in 3D but this time by:

1) using the space of moments (*p*) wherein Hesienberg's relation can be written with respect to a specific direction, for example *x* so that $\Delta x \cdot \Delta p_x = h$; and

2) in *k* space where the use of Heisenberg's relation can be made and justified.

Answers

1) In 3D, Heisenberg's equation is written: $\Delta x\ \Delta y\ \Delta z\ \Delta p_x\ \Delta p_y\ \Delta p_z = h^3$, so that by making $V = \Delta x\ \Delta y\ \Delta z$ (space within which the particle can move) we have:

$$\Delta p_x\ \Delta p_y\ \Delta p_z = \frac{h^3}{V}.$$

A free electron placed in a potential of zero exhibits a relation between *E* and *p* that is: $E = \frac{p^2}{2m}$. Quantum states for energies between *E* and *E* + *dE* in moment space are situated between spheres with radii $p\ (= \sqrt{2\text{mE}})$ and *p* + *dp*. The volume between two spheres is $4\pi p^2 dp$. This volume can hold a number of cells of volume $\Delta p^3 = \Delta p_x \Delta p_y \Delta p_z$ given by $\frac{4\pi p^2 dp}{h^3/V}$.

By making *V* = 1 (to obtain a given density) and recognizing that we can place two electrons per cell, the state density function *Z*(*E*) is such that $Z(E)dE = 2 \cdot \frac{4\pi p^2 dp}{h^3}$. The differentiation of $p^2 = 2mE$ gives $2pdp = 2mdE$, from which $2p^2dp = (2m)^{3/2} E^{1/2} dE$, and we directly obtain:

$$Z(E)dE = 4\pi \left(\frac{2m}{h^2}\right)^{3/2} E^{1/2} dE.$$

2) With $p = \hbar k$, we have $\Delta p_x = \hbar \Delta k_x$, from which $\Delta x \cdot \Delta k_x = 2\pi$ (Heisenberg's other equation!). From this we deduce that $\Delta k_x \Delta k_y \Delta k_z = \frac{(2\pi)^3}{V}$. This relation is identical to that deduced for the quantification with progressive boundary conditions (PBC), where for a free electron we have $\psi_k(x) = Ae^{ik_x x} = \psi_k(x+L) = Ae^{ik_x(x+L)}$, from which $e^{ik_x L} = 1$, so that $k_x L = 2\pi(n)$, so that between adjacent values of n, k_x varies by $\Delta k_x = \frac{2\pi}{L}$. This result makes it possible to justify Heisenberg's equation, as L is the space Δx in which the electron can probably be found.

In k space, the volume for energies between E and $E + dE$ is situated between the spheres of radii k $(= \sqrt{2\mathrm{mE}}/\hbar)$ and $k + dk$, equal to a volume of $4\pi k^2 dk$. Into this volume we can place $\frac{4\pi k^2 dk}{(8\pi^3)/V}$ cells, so that to obtain the state density function when $V = 1$, we have $Z(E)dE = \frac{k^2 dk}{\pi^2}$ electrons. With $k^2 = \frac{2mE}{\hbar^2}$, we can state that $kdk = \frac{m\,dE}{\hbar^2}$, and we once again obtain:

$$Z(E)dE = 4\pi\left(\frac{2m}{h^2}\right)^{3/2} E^{1/2} dE.$$

2.7.6. *Problem 6: general equations for the state density functions expressed in reciprocal (k) space or in energy space*

1) Based on the reasoning given in problem 5 question 2, show how it is possible to write the 3D relation between the state density function in energy [$Z(E)$] and the state density function in k space [$n(k)$] in the form:

$$Z(E)dE = \int_{\delta V(E)} n(k)d^3k,$$ detailing the significance of $\delta V(E)$.

2) For theories more refined than that based on the free electron, the dispersion equation for $E(k)$ can be quite complex, making the determination of $\delta V(E)$ more

difficult. Show how the following general equation for 3D space can be derived and detail the significance of *S*(*E*).

$$Z(E) = \frac{1}{4\pi^3} \int_{S(E)} \frac{dS}{\left|\overrightarrow{grad}_k E\right|}.$$

3) What happens to the preceding equation in 2D? Apply it to free electrons to obtain the equation for *Z*(*E*) as found in question 2 of the preceding problem 3.

Answers

1) By definition, the density function of energy states *Z*(*E*) is such that the number of states with energies between *E* and *E* + *dE* is given by *Z*(*E*)*dE*. For free electrons (where $E = \frac{\hbar^2 k^2}{2m}$, so that $k = \frac{\sqrt{2mE}}{\hbar}$), the corresponding energy volume is situated in *k* space between two spheres of radii *k* and *k* + *dk*, that is, of volume $dV_k = 4\pi k^2 dk$. The transition from the function *n*(*k*), i.e. the state density in *k* space *k*, to the function *Z*(*E*) – the density of state in energy space – is thus given via the equation:

$$Z(E)dE = n(k) 4\pi k^2 dk.$$

Given that $n(k) = \frac{1}{4\pi^3}$ (see equation [2.27]), we once again find the equation in problem 5, i.e. $Z(E)dE = \frac{k^2 dk}{\pi^2}$.

For free electrons, as *n*(*k*) is a constant in *k* space (equation [2.27]) we can thus write that $Z(E)dE = n(k)\int_{\delta V(E)} d^3k$ where $\int_{\delta V(E)} d^3k = 4\pi k^2 dk$.

$\delta V(E)$ thus represents the volume of reciprocal space between the surface *S*(*E*) of the equi-energy *E* and the surface *S*(*E* + *dE*) of the equi-energy *E* + *dE*.

In more general terms, if *n*(*k*) varies with k, the preceding relationship should thus be written as:

$$Z(E)dE = \int_{\delta V(E)} n(k) d^3k.$$

2) The elemental volume d^3k can be seen as the product of elemental surface d^2S carried through area $S(E)$, in other words the surface of the equi-energy E in reciprocal space through the space between areas $S(E)$ and $S(E + dE)$ along a line normal to $S(E)$. This normal $S(E)$ is collinear to $\overrightarrow{\text{grad}}_k E$ so that we can state that:

$$d^3k = d^2S \cdot \left|\overrightarrow{\text{grad}}_E\, k\right| dE = d^2S \, \frac{dE}{\left|\overrightarrow{\text{grad}}_k\, E\right|}.$$

Placing this in $Z(E)dE = \int_{\delta V(E)} n(k)d^3k$ where $n(k) = \frac{1}{4\pi^3}$ (equation independent from the dispersion equation), we have $Z(E)dE = \frac{1}{4\pi^3}\int_{\delta V(E)} \frac{d^2S}{\left|\overrightarrow{\text{grad}}_k\, E\right|}\, dE$, from which:

$$Z(E) = \frac{1}{4\pi^3} \int_{S(E)} \frac{d^2S}{\left|\overrightarrow{grad}_k\, E\right|}$$

where $S(E)$ is the surface of the equi-energy E in reciprocal space.

NOTE.– We can quickly verify the veracity of equation [2.28] for free electrons. Here, in effect we find:

$$\begin{aligned} Z(E) &= \frac{1}{4\pi^3} \int_{S(E)} \frac{d^2S}{\left|\overrightarrow{\text{grad}}_k\, E\right|} = \frac{1}{4\pi^3} \int_{S(E)} \frac{d^2S}{\frac{d(\hbar^2k^2/2m)}{dk}} \\ &= \frac{1}{4\pi^3} \int_{S(E)} \frac{d^2S}{\frac{\hbar^2 k}{m}} = \frac{1}{4\pi^3} \frac{4\pi k^2}{\frac{\hbar^2 k}{m}} = \frac{4mk}{h^2} = \frac{4m}{h^2}\frac{\sqrt{2mE}}{\hbar} = \frac{4\pi}{h^3}(2m)^{3/2}\sqrt{E}. \end{aligned}$$

3) In 2D, $n(k)$ must be such that $n(k)\left[\Delta k^2\right]_{S=1} = 2$, so that with $\left[\Delta k^2\right]_{S=1} = \left[\frac{2\pi}{L_1}\frac{2\pi}{L_2}\right]_{S=1} = \left[\frac{4\pi^2}{S}\right]_{S=1} = 4\pi^2$ we have $n(k) = \frac{1}{2\pi^2}$. Analogously to

question 2, and on accepting that the equi-energy surface S(E) becomes an equi-energy line *L*(*E*) in 2D, we have:

$$Z(E) = \frac{1}{2\pi^2} \int_{L(E)} \frac{dL}{\left|\overrightarrow{\mathrm{grad}}_k\, E\right|}.$$

Its application to free electrons makes it possible to again find the equation for *Z*(*E*) previously given in the answer to problem 3. In effect, we now have:

$$Z(E) = \frac{1}{2\pi^2} \int_{L(E)} \frac{dL}{\left|\overrightarrow{\mathrm{grad}}_k\, E\right|} = \frac{1}{2\pi^2} \frac{2\pi k}{\frac{\hbar^2 k}{m}} = \frac{m}{\pi\hbar^2} = \text{constant}.$$

Chapter 3

The Origin of Band Structures within the Weak Band Approximation

3.1. Bloch function

As presented in Chapter 1, the weak bond applies for electrons placed in a periodic potential, as represented in Figure 3.1.

3.1.1. *Introduction: effect of a cosinusoidal lattice potential*

For a free electron – in a zero order approximation – the potential is that of a flat-bottomed bowl (as indicated by the line through the nuclei in Figure 3.1). For this system, already studied in Chapter 2, the potential (potential energy) is such that $V = V^0 = 0$, and the Schrödinger amplitude equation is:

$$\Delta\psi^0 + \frac{2m}{\hbar^2} E\,\psi^0 = 0.$$

With this and by making $\mathrm{k}^2 = \frac{2\mathrm{m}}{\hbar^2}\mathrm{E}$, it is now possible to write solutions in the form $\psi^0 = e^{ikx}$ for the wave function, and $E = E^0 = \frac{\hbar^2}{2m} k^2$ for the energy. Thus, at this level of the zero order approximation we have a parabolic form representing $E = f(k)$.

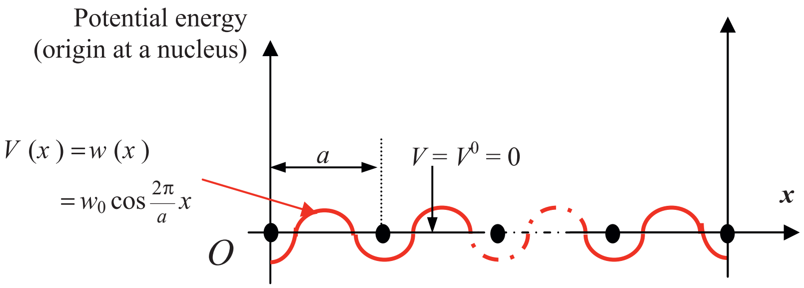


Figure 3.1. *Curve of the potential energy* $w(x) = w_0 \cos \frac{2\pi}{a} x$, *showing that* $w_0 < 0$

The perturbation of potential by the lattice generates a periodic potential, as we have seen in Chapter 1. The first term of the Fourier series gives a first approximation, as in $V \approx V^{(1)} = w(x) = w_0 \cos \frac{2\pi}{a} x$. The wave function is thus also perturbed. It takes on the form of a Bloch function which, as detailed below, is of the form $\psi_k(x) = \psi^0 u(x) = e^{ikx} u(x)$, where $u(x)$ is a periodic function (with a period equal to that of the lattice).

3.1.2. *Properties of a Hamiltonian of a semi-free electron*

In the approximation of a semi-free electron, the equation in proper terms, $H\psi = E\psi$, is such that with $V = w_0 \cos \frac{2\pi}{a} x$ we have:

$$-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} + V\psi = E\psi,$$

so that in addition:

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar}(E - V)\psi = 0.$$

The semi-free electron's Hamiltonian is therefore:

$$H = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \qquad \text{with } V(x) = V(x + a).$$

We can thus deduce that $H(x) = H(x+a)$ as $V(x) = V(x+a)$ and $\frac{d^2}{dx^2} = \frac{d^2}{d(x+a)^2}$ because, in simple terms, we have $\frac{d}{dx} = \frac{d}{d(x+a)}\frac{d(x+a)}{dx} = \frac{d}{d(x+a)}$.

We can consequently conclude that a semi-free electron's Hamiltonian is invariant with respect to a translation (T_a) of modulus a.

Using the properties established in Chapter 1, the search for the proper Hamiltonian functions associated with a semi-free electron can in fact lead us to search for the proper functions of the translational operator T_a.

3.1.3. *The form of proper functions*

We will denote the proper functions of translation operator T_a as $\psi_k(x)$.

They will be such that:

$$T_a \psi_k(x) = c_{k,a} \psi_k(x)$$

where $c_{k,a}$ is a proper value of operator T_a.

Depending on the definition of operator T_a, we also have:

$$T_a \psi_k(x) = \psi_k(x+a).$$

Now, let us try and obtain the precise form of the coefficients $c_{k,a}$. In order to do this, we will apply two successive translations of modulus a_1 and then a_2 to the function $\psi_k(x)$.

In this example we will use $a_1 = a$ and $a_2 = -a$ so that the application of operator $T_{a_2} \circ T_{a_1}$ results in a return to the departure point and thus leaves function $\psi_k(x)$ unchanged. We can thus write that:

$$T_{a_1}\psi_k(x) = c_{k,a}\psi_k(x) = \psi_k(x+a)$$

$$\left.\begin{array}{r}\left(T_{a_2} \circ T_{a_1}\right)\psi_k(x) = T_{a_2}(T_{a_1}\psi_k(x)) = T_{a_2}\left(c_{k,a}.\psi_k(x)\right) = c_{k,-a}c_{k,a}\psi_k(x) \\ T_{a_2}(T_{a_1}\psi_k(x)) = \psi_k(x+a-a) = \psi_k(x) \qquad\qquad\qquad\qquad\end{array}\right\}$$

$$\Rightarrow c_{k,-a}c_{k,a} = 1.$$

$$\begin{tikzcd}
& T_{a_1}\psi_k(x)=\psi_k(x+a_1) \arrow[dr, "T_{a_2}"] & \\
\psi_k(x) \arrow[ur, "T_{a_1}"] \arrow[rr, "T_{a_2}\circ T_{a_1}"'] & & \left(T_{a_2}\circ T_{a_1}\right)\psi_k(x)=\psi_k(x+a_1+a_2)
\end{tikzcd}$$

An acceptable solution is $c_{k,a}=e^{ika}$ as it is in accordance with the preceding condition as $e^{-ika}e^{ika}=1$.

Finally, the equation for the proper values can be written as:

$$T_a\psi_k(x)=e^{ika}\psi_k(x).$$

Now, let us determine the form of the proper function $\psi_k(x)$ of a semi-free electron (in first-order approximation). In order to do this, we can first of all note that the zero-order solution is of the form $\psi_k^0(x)=e^{ikx}$. To reach the first-order approximation (where the function is denoted by $\psi_k^{(1)}(x)\equiv\psi_k(x)$) the solution to the function of the zero order must be perturbed by a function $u(x)$ such that: $\psi_k^{(1)}(x)\equiv\psi_k(x)=\psi_k^0(\mathrm{x})\,u(x)=e^{ikx}u(x)$.

Now we need to find the form that gives u(x). To do that, we look at the proper functions $\psi_k(x)$ that can be chosen in the form $\psi_k(x)=e^{ikx}u(x)$ with the condition that $u(x) = u(x+a)$. These functions are called Bloch functions. In effect, we have:

$$\left.\begin{array}{l}
T_a\psi_k(x)=c_{k,a}\psi_k(x) \overset{\psi_k(x)=e^{ikx}u(x)}{=} c_{k,a}\mathrm{e}^{\mathrm{ikx}}\mathrm{u(x)} \overset{\mathrm{c}_{\mathrm{k,a}}=e^{ika}}{=} e^{ika}e^{ikx}u(x)\\
\text{and}\\
T_a\psi_k(x)=\psi_k(x+a) \overset{\left[\psi_k(x)\right]_{"x=x+a"}=\left[e^{ikx}u(x)\right]_{"x=x+a"}}{=} e^{ik(x+a)}u(x+a)
\end{array}\right\}\Rightarrow$$

$$u(x)=u(x+a).$$

To conclude, the proper functions for the semi-free electron are in the form $\psi_k(x) = e^{ikx}u(x)$ and are called Bloch functions. Function $u(x)$ is periodic with a period equal to a, the lattice repeat unit.

3.2. Mathieu's equation

3.2.1. *Form of Mathieu's equation*

The form of this equation is that of the Schrödinger equation where $V = w_0 \cos\frac{2\pi}{a}x$. By denoting the mass of the electron as μ we thus have the following Mathieu's equation:

$$\frac{d^2\psi}{dx^2} + \frac{2\mu}{\hbar^2}\left(E - w_0\cos\frac{2\pi}{a}x\right)\psi = 0. \qquad [3.1]$$

3.2.2. *Wave function in accordance with Mathieu's equation*

Following on from Bloch's theorem, Mathieu's equation should be written $\psi_k(x) = e^{ikx}u(x)$, where $u(x)$ is a periodic function of period a, the lattice repeat unit. On introducing the angular frequency ω that is tied to the repeat unit of the lattice by equation $a = \frac{2\pi}{\omega}$, it is possible to develop the periodic function $u(x)$ as a Fourier series, such that:

$$u(x) = \sum_n A_n e^{-in\omega x}$$

where n is an integer and $A_n = \frac{\omega}{2\pi}\int_{\frac{2\pi}{\omega}} u(x)e^{-in\omega x}dx$.

To within a normalization factor (A_0), the function $\psi_k(x)$ must therefore be in the form:

$$\psi_k(x) = e^{ikx}\left[1 + \sum_{n\neq 0} A_n e^{-in\omega x}\right]. \qquad [3.2]$$

In equation [3.2], the first term (associated with unit 1 of the square bracket) corresponds to the zero-order wave function of an electron ($\psi_k^0(x) \propto e^{ikx}$), and can perhaps be seen as the principal part of this wave function. As a consequence it is possible to write (still within the normalization term) that:

$$\psi_k(x) = e^{ikx} + \sum_{n \neq 0} A_n e^{i(k - \omega n)x} . \qquad [3.3]$$

From this we can deduce that:

$$\begin{aligned} \frac{\partial \psi_k}{\partial x} &= ike^{ikx} + \sum_{n \neq 0} i(k - n\omega)A_n e^{i(k - n\omega)x} \\ \frac{\partial^2 \psi_k}{\partial x^2} &= -k^2 e^{ikx} - \sum_{n \neq 0} (k - n\omega)^2 A_n e^{i(k - n\omega)x} . \end{aligned} \qquad [3.4]$$

The solution for zero order energy E, written as $E^0 = \frac{\hbar^2 k^2}{2m}$, can be introduced into equation [3.1], so that we then obtain the energy E as a precise function of E^0 and bring in the term $[E - E^0]$ (which is a small term for the first order term). This gives:

$$\frac{d^2\psi}{dx^2} + \left[\frac{2\mu}{\hbar^2}(E - E^0 - w_0 \cos \omega x) + k^2\right]\psi = 0. \qquad [3.5]$$

Substituting equations [3.3] and [3.4] into equation [3.5] gives terms in $k^2 e^{ikx}$ that cancel out:

$$\frac{d^2\psi}{dx^2} + k^2\psi + \frac{2\mu}{\hbar^2}\left[E - E^0 - w_0 \cos \omega x\right]\psi = 0,\ k^2 e^{ikx}$$

and hence we can now write that:

$$\begin{aligned} &\sum_{n \neq 0} A_n e^{i(k - n\omega)x}\left[k^2 - (k - n\omega)^2\right] + \frac{2\mu}{\hbar^2}\left[E - E^0 - w_0 \cos \omega x\right]e^{ikx} \\ &\quad + \frac{2\mu}{\hbar^2} \sum_{n \neq 0}\left[E - E_0 - w_0 \cos \omega x\right]A_n e^{i(k - n\omega)x} = 0. \end{aligned} \qquad [3.6]$$

In equation [3.6], the third term brings in the coefficients $(E - E^0)A_n$ and $w_0 A_n$, which are second-order terms, as $(E - E^0)$, w_0 and $[A_n]_{n \neq 0}$ are all small terms of the first order that can be neglected in an approximate calculation. This is because to zero order, the energy is E^0, the potential energy is $V = V^0 = 0$, and the development of wave function is constrained to the single term for A_0, so that $[A_n]_{n \neq 0} \simeq 0$.

Mathieu's equation is now reduced here to:

$$\sum_{n \neq 0} A_n e^{i(k - n\omega)x} \left[k^2 - [k - n\omega]^2\right] + \frac{2\mu}{\hbar^2}\left[E - E^0 - w_0 \cos \omega x\right] e^{ikx} = 0. \quad [3.7]$$

The resolution method, i.e. to obtain coefficient A_m (of the development of ψ), is to multiply equation [3.7] by $e^{-i(k - m\omega)x}$ and then integrate over a repeat unit, in other words from 0 to a.

Thus, the first step, multiplication by $e^{-i(k - m\omega)x}$, gives us:

$$\sum_{n \neq 0} A_n \left[k^2 - (k - n\omega)^2\right] e^{i\omega x(m - n)} + \frac{2\mu}{\hbar^2}\left[E - E^0 - w_0 \cos \omega x\right] e^{im\omega x} = 0.$$

Here though we should note that:

– the only term of the first sum which does not cancel itself out on integration is when $m = n$, and corresponds in the summation to the coefficient A_m. The value of this term after integration is:

$$\begin{aligned}\int_0^a A_m \left[k^2 - (k - m\omega)^2\right] dx &= A_m \left[k^2 - (k - m\omega)^2\right] a \\ &= aA_m \left[k^2 - k^2 - m^2\omega^2 + 2km\omega\right] \\ &= 2m\omega a A_m \left[k - \frac{m\omega}{2}\right];\end{aligned}$$

– $\int_0^a (E - E^0) e^{im\omega x} dx = 0$;

– the second term will be different from zero only for terms in cos² that will be obtained when $m = \pm 1$.

Thus, we can conclude that except for A_1 and A_{-1} the coefficients of A_m are all zero.

When $m = 1$, we have (noting that $\int_0^a \cos\omega x e^{i\omega x} dx = \int_0^a \cos^2\omega x dx + i \underbrace{\int_0^a \cos\omega x \sin\omega x dx}_{=0} = \frac{a}{2}$):

$$2\omega a A_1 \left(k - \frac{\omega}{2}\right) = \frac{2\mu}{\hbar^2} w_0 \int_0^a \cos\omega x e^{i\omega x} dx = \frac{\mu}{\hbar^2} w_0 a$$
$$\Rightarrow A_1 = \frac{\mu w_0}{2\hbar^2 \omega \left(k - \frac{\omega}{2}\right)}. \qquad [3.8]$$

Similarly when $m = -1$ we have:

$$-2\omega a A_{-1} \left(k + \frac{\omega}{2}\right) = \frac{2\mu}{\hbar^2} w_0 \int_0^a \cos\omega x e^{-i\omega x} dx$$
$$\Rightarrow A_{-1} = -\frac{\mu w_0}{2\hbar^2 \omega \left(k + \frac{\omega}{2}\right)}. \qquad [3.9]$$

Finally, within a first-order approximation, the form of equation [3.2] for the wave function is now limited to:

$$\psi = e^{ikx} \left[1 + \frac{\mu w_0}{2\hbar^2\omega} \left(\frac{e^{-i\omega x}}{k - \frac{\omega}{2}} - \frac{e^{i\omega x}}{k + \frac{\omega}{2}} \right) \right]. \qquad [3.10]$$

3.2.3. *Energy calculation*

Here we are seeking to determine the term $(E - E^0)$ introduced into equation [3.6]. The expression is multiplied by e^{-ikx} and then integrated from 0 to a. Stepwise, we obtain the initial multiplication by e^{-ikx} which gives

$$\sum_{n\neq 0} A_n e^{-in\omega x}\left[k^2-(k-n\omega)^2\right]+\frac{2\mu}{\hbar^2}\left[E-E^0-w_0\cos\omega x\right]$$
$$+\frac{2\mu}{\hbar^2}\sum_{n\neq 0}\left[E-E^0-w_0\cos\omega x\right]A_n e^{-in\omega x}=0$$

– The first sum integrated is zero.

– The second integrated term gives $\frac{2\mu}{\hbar^2}(E-E^0)a$ (and now we realize that if we had restrained equation [3.6] to [3.7] we would have simply found $E=E^0$, which would not have given E in this new approximation).

– As only A_1 and A_{-1} are different at zero, the third and final sum gives non-zero terms:

$$-\frac{2\mu}{\hbar^2}w_0\int_0^a\cos\omega x\left[A_1e^{-i\omega x}+A_{-1}e^{i\omega x}\right]dx=-\frac{\mu w_0 a}{\hbar^2}(A_1+A_{-1})$$
$$=-\frac{\mu^2 w_0^{\,2}a}{2\hbar^4\omega}\left(\frac{1}{k-\frac{\omega}{2}}-\frac{1}{k+\frac{\omega}{2}}\right)$$
$$=-\frac{\mu^2 w_0^{\,2}a\omega}{2\hbar^4\omega\left(k^2-\frac{\omega^2}{4}\right)}.$$

We finally obtain:

$$\frac{2\mu}{\hbar^2}a(E-E^0)=\frac{\mu^2 w_0^{\,2}a}{2\hbar^4\left(k^2-\frac{\omega^2}{4}\right)}\Rightarrow E-E^0=\frac{\mu w_0^{\,2}}{4\hbar^4\left(k^2-\frac{\omega^2}{4}\right)},\qquad [3.11]$$

so that with $a = \frac{2\pi}{\omega}$:

$$E = E^0 + \frac{\mu w_0{}^2}{4\hbar^2\left(k^2 - \frac{\pi^2}{a^2}\right)}. \qquad [3.12]$$

When deducing equation [3.6] from equation [3.7] to determine the coefficient A_n, and coefficients A_1 and A_{-1} (the latter are also used in the determination of $E - E^0$), we made the assumption that the terms for A_n were small (and importantly *a posteriori* A_1 and A_{-1}) much as $[E - E^0]$. The results confirm that the assumption was reasonable except when k is close to $\pm\frac{\pi}{a}$. When $k \simeq \pm\frac{\pi}{a}$ we can see that $[E - E^0]$ (in equation [3.11]) and A_1 and A_{-1} (equations [3.8] and [3.9]) become very large. A direct calculation is thus required for these values when $k \simeq \pm\frac{\pi}{a}$. We will now look at the calculation of energy when $k \simeq \pm\frac{\pi}{a}$.

3.2.4. *Direct calculation of energy when* $k \simeq \pm\frac{\pi}{a}$

In this calculation we will limit ourselves to the case of $k = \frac{\pi}{a}$. The other case, $k = -\frac{\pi}{a}$, leads to a similar result. Following from the preceding example (equation [3.5]), we can see that when $k = \frac{\pi}{a}$, only coefficient A_1 is large: the other A_n coefficients remain small. Hence, the only approximation made this time outside of what remains a direct calculation gives us a new form of equation [3.3] (still within a normalization factor) for the wave function:

$$\psi_k(x) = e^{ikx} + A_1 e^{i(k-\omega)x}. \qquad [3.13]$$

With $\frac{d^2\psi_k}{dx^2} = -k^2 e^{ikx} - (k-\omega)^2 A_1 e^{i(k-\omega)x}$, Mathieu's equation [3.1] can be written as:

$$-k^2 e^{ikx} - (k-\omega)^2 A_1 e^{i(k-\omega)x} + \frac{2\mu}{\hbar^2}(E - w_0 \cos\omega x)e^{ikx} + \frac{2\mu}{\hbar^2} A_1 (E - w_0 \cos\omega x) e^{i(k-\omega)x} = 0.$$

This equation can be rearranged in the form:

$$\left[\frac{2\mu}{\hbar^2}(E - w_0 \cos\omega x) - k^2\right] e^{ikx} + \left[\frac{2\mu}{\hbar^2} A_1 (E - w_0 \cos\omega x) - A_1 (k-\omega)^2\right] e^{i(k-\omega)x} = 0.$$

To resolve this equation and determine the energy, we again use the general principles that were applied to Mathieu's equation. That is, we successively multiply the equation by e^{-ikx} and then by $e^{-i(k-\omega)x}$ (the conjugated terms of the imaginary exponentials in Mathieu's equation), each time integrating the obtained equations over the range 0 to a. Thus, the two following equations are obtained:

$$\begin{cases} \left(\frac{2\mu E}{\hbar^2} - k^2\right) a - A_1 \frac{\mu w_0}{\hbar^2} a = 0 & [3.14] \\ -\frac{\mu w_0}{\hbar^2} a + A_1 \left[\frac{2\mu E}{\hbar^2} - (k-\omega)^2\right] a = 0. & [3.15] \end{cases}$$

In order to ensure that these two equations are compatible, the determinant of the system must be equal to zero. This gives the following relation (which can also be obtained by eliminating A_1 from equations [3.14] and [3.15]):

$$\left(\frac{2\mu E}{\hbar^2} - k^2\right)\left(\frac{2\mu E}{\hbar^2} - [k-\omega]^2\right) = \frac{\mu^2 w_0^2}{\hbar^4}. \qquad [3.16]$$

On multiplying the two members by $\frac{\hbar^4}{4\mu^2}$, we obtain:

$$\left(E - \frac{\hbar^2 k^2}{2\mu}\right)\left(E - \frac{\hbar^2 [k - \omega]^2}{2\mu}\right) = \frac{w_0^2}{4}, \qquad [3.17]$$

so that when $k = \frac{\pi}{a}$ (and using $\omega = \frac{2\pi}{a}$) we have

$$\left[E - \frac{\hbar^2 \left(\frac{\pi}{a}\right)^2}{2\mu}\right]^2 = \frac{w_0^2}{4},$$

from which we find:

$$[E]_{k=\frac{\pi}{a}} = \frac{\hbar^2 \left(\frac{\pi}{a}\right)^2}{2\mu} \pm \frac{w_0}{2}. \qquad [3.18]$$

NOTE.– When $k = -\frac{\pi}{a}$, a similar set of calculations will yield the same energy values as found in equation [3.18].

3.3. The band structure

3.3.1. *Representing E = f (k) for a free electron: a reminder*

For a free electron (in a zero-order approximation), the potential is considered to be like a flat-bottomed bowl ($V = V^0 = 0$ where a horizontal line passes through the nucleus depicted in Figure 3.1). As shown in Chapter 2, the progressive solutions for this system are in the form $\psi = e^{ikx}$ (wave function) and $E = E^0 = \frac{\hbar^2}{2m} k^2$ (energy). $E = f(k)$ is thus shown as being parabolic at a zero-order approximation.

3.3.2. *Effect of a cosinusoidal lattice potential on the form of wave function and energy*

3.3.2.1. *When k is very different from* $\pm\frac{\pi}{a}$

We have seen above how the wave function is given to the first order by equation [3.10] and corresponds to a progressive wave. For its part, the energy is given by equation [3.12], as in: $E = E^0 + \frac{\mu {w_0}^2}{4\hbar^2\left(k^2 - \frac{\pi^2}{a^2}\right)}$. This curve is taken away from the parabolic form shown in Figure 3.2 by the presence of the term $\frac{\mu {w_0}^2}{4\hbar^2\left(k^2 - \frac{\pi^2}{a^2}\right)}$.

3.3.2.2. *When* $k \approx \pm\frac{\pi}{a}$

3.3.2.2.1. Evolution of energy

Noting that the expression for energy obtained at zero order is $E^0 = \frac{\hbar^2 k^2}{2\mu}$, we can insert this into equation [3.12] to obtain an energy equation for when $k = \pm\frac{\pi}{a}$:

$$[E]_{k=\frac{\pi}{a}} = \left[E^0\right]_{k=\frac{\pi}{a}} \pm \frac{w_0}{2}. \qquad [3.16]$$

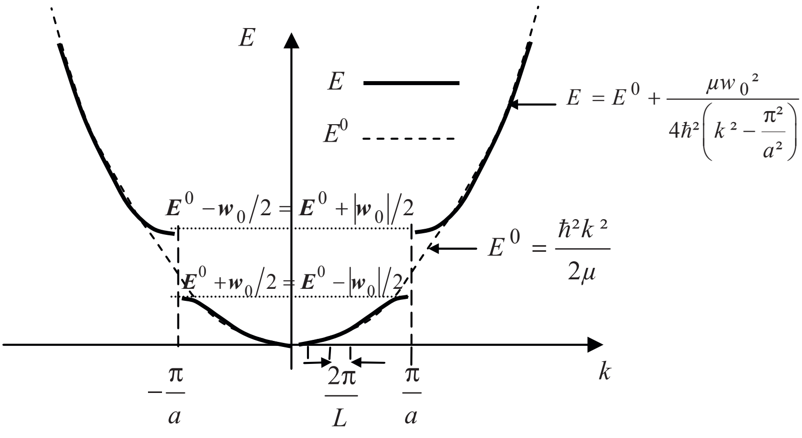


Figure 3.2. *Curve E = f(k)*

We can thus see in Figure 3.2 (where k values are quantified according to $k_n = n\,2\pi/L$) that as k tends towards $\frac{\pi}{a}$, the energy moves away from the value given in equation [3.8] and tends towards:

$$[E]_{k=\frac{\pi}{a}} = [E^0]_{k=\frac{\pi}{a}} - \frac{|w_0|}{2}.$$

In algebraic terms, as $w_0 < 0$ (see Figure 3.1), we should write that E tends towards:

$$[E]_{k=\frac{\pi}{a}} = [E^0]_{k=\frac{\pi}{a}} + \frac{w_0}{2} = \frac{\hbar^2\left(\frac{\pi^2}{a^2}\right)}{2\mu} + \frac{w_0}{2}.$$

As k increases towards the value $\frac{\pi}{a}$, energy passes abruptly to the value given by:

$$[E]_{k=\frac{\pi}{a}} = [E^0]_{k=\frac{\pi}{a}} + \frac{|w_0|}{2} = \frac{\hbar^2\left(\frac{\pi^2}{a^2}\right)}{2\mu} + \frac{|w_0|}{2},$$

and then increases so that it closes in on the curve given by the relation:

$$E = E^0 + \frac{\mu w_0^{\,2}}{4\hbar^2\left(k^2 - \frac{\pi^2}{a^2}\right)}.$$

3.3.2.2.2. Stationary form of wave functions when $k = \pm\frac{\pi}{a}$

When $k = \frac{\pi}{a}$, the wave function given by equation [3.9] is in fact in the form:

$$\psi_k(x) = A_0 e^{ikx} + A_1 e^{i(k-\omega)x} = A_0 e^{i\frac{\pi}{a}x} + A_1 e^{i\left(\frac{\pi}{a}-\frac{2\pi}{a}\right)x} = A_0 e^{i\frac{\pi}{a}x} + A_1 e^{-i\frac{\pi}{a}x}$$

where $\omega = \frac{2\pi}{a}$ and A_0 is introduced as a normalization factor.

In fact, the function for the incident wave (amplitude term $A_0 e^{i\frac{\pi}{a}x}$) and the reflected wave (amplitude $A_1 e^{-i\frac{\pi}{a}x}$) is a stationary solution (see also, for example, Chapter 6 [MOL 07b]). Found by calculation, this result can also be determined using more direct physics (see section 3.4), where a greater understanding of the forbidden bands is possible. This is because the two stationary solutions (in terms of cosine and sine in section 2.2.2) – and each for a solution in energy – define the values in which a forbidden band resides. Their energy difference is equal to the width of the forbidden band (or "forbidden energy") of the material.

3.3.3. *Generalization: effect of a periodic non-ideally cosinusoidal potential*

In the most general of terms, the potential interacting with the lattice must be seen as simply periodic (with the period being that of the lattice repeat unit). It has a form $P(x)$ that is non-ideally cosinusoidal, which up till now has been introduced using the function $V(x) = w_0 \cos\frac{2\pi}{a}x$.

Equation [3.1] can thus be written as:

$$\frac{d^2\psi}{dx^2} + \frac{2\mu}{\hbar^2}(E - P(x))\psi = 0,$$

and as such corresponds to the general form of Mathieu's equations. The general method to resolve these equations is similar to that shown in section 3.2. The periodic potential $P(x)$ is thus treated as a Fourier series that brings in the terms $\cos n\omega x = \cos 2\pi n\frac{x}{a}$. It is thus possible to find that in the wave function expression, coefficients A_n and A_{-n} are non-zero and not small when k is close to $\pm\frac{n\pi}{a}$. These discontinuities, which decrease as n increases, occur for energy values found around $n^2\frac{\hbar^2\left(\frac{\pi^2}{a^2}\right)}{2\mu}$.

We thus obtain a representation of the energy bands, as shown in Figure 3.3.

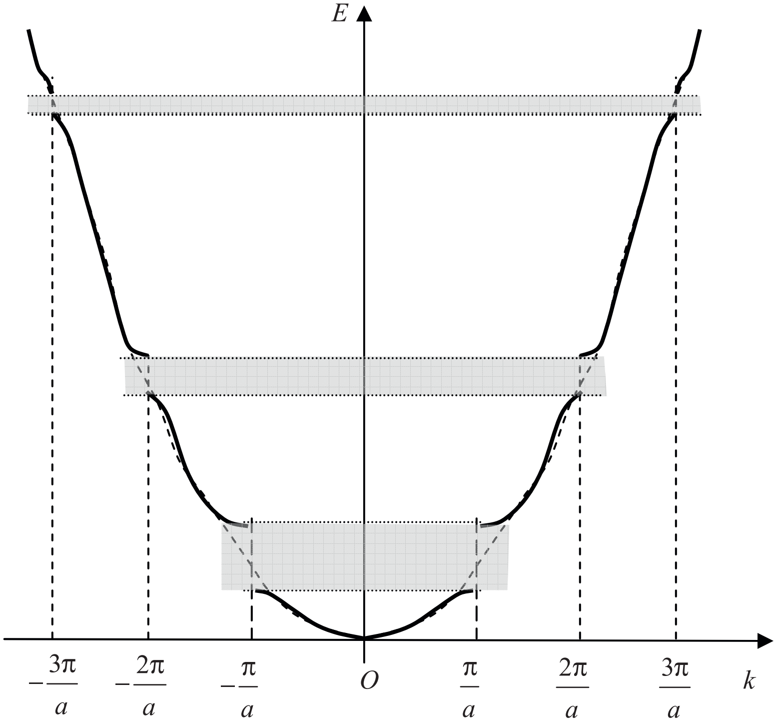


Figure 3.3. *E = f(k) with discontinuities when* $k = \pm n\frac{\pi}{a}$ *that decrease as n increases*

3.4. Alternative presentation of the origin of band systems via the perturbation method

3.4.1. *Problem treated by the perturbation method*

For a weak bond, the perturbation potential $V(x) = w_0 \cos\frac{2\pi}{a}x$ can be assumed to be small (the amplitude w_0 is small) and hence the perturbation method widely used in quantum mechanics can be applied. If ψ^0 represents the wave function of the non-perturbed state characterized by the energy E^0, then the energy of the system perturbed by the Hamiltonian perturbation H_{pert} ($\equiv V(x)$ in this case) is given by:

$$E = E^0 + E_{\text{pert}} \quad \text{where} \quad E_{\text{pert}} = \Delta E = \left\langle \psi^0 | \mathrm{H}_{\text{pert}} | \psi^0 \right\rangle.$$

3.4.2. *Physical origin of forbidden bands*

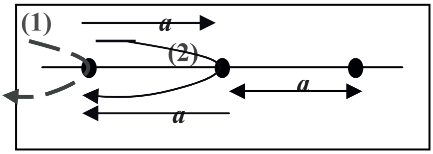


Figure 3.4. *Bragg reflection for a 1D crystal*

Figure 3.4 shows a periodic chain subjected to an incident ray, which is reflected by the atoms of the lattice. There is an additive wave interference following reflection if the difference in step (Δ) between two waves is equal to a whole multiple of the incoming waves wavelength. For a 1D system, the value of Δ between wave 1 and wave 2 following reflection is given by $\Delta = 2a$ so that only incident waves of wavelength (λ_n), such that $\Delta = 2a = n\lambda_n$, will give a maximum reflection. With the wave vector module being $k = \frac{2\pi}{\lambda}$, the incident waves which undergo the maximum reflection (otherwise known as diffraction) are also such that:

$$k = k_n = \frac{2\pi}{\lambda_n} = n\frac{\pi}{a}. \qquad [3.19]$$

In other words, they satisfy the Bragg condition.

For a weak bond, we can assume that the incident wave associated with an electron in the bond is only weakly perturbed by the linear chain and that its amplitude can be written using a zero-order approximation, that is in the form:

$$\psi_k^0 = Ae^{ikx}.$$

The time-dependent incident wave is thus $[\Psi_k^0(x,t)]_{\text{inc.}} = Ae^{i(kx-\omega t)}$. This is the expression for an incident plane progressive wave moving towards $x > 0$.

When the equation $k = k_n$ is precisely fulfilled, then this incident wave $[\Psi_{k_n}^0(x,t)]_{\text{inc.}} = Ae^{i(k_n x-\omega t)}$ is reflected as a wave propagating towards $x < 0$ and is given by $[\Psi_{k_n}^0(x,t)]_{\text{refl.}} = Ae^{i(k_n x+\omega t)}$. The superposition of these two types of waves (incident and reflected) establishes a stationary wave regime. This regime has two solutions (symmetric and asymmetric, as established in section 2.2.2) that can be written, to within the proximity of a normalization coefficient, as follows:

$$\Psi^+ \propto \left[e^{+i\frac{n\pi}{a}x} + e^{-i\frac{n\pi}{a}x} \right] e^{-i\omega t} \approx \cos\left(\frac{n\pi}{a}x\right) e^{-i\omega t} \qquad [3.20]$$

and

$$\Psi^- \propto \left[e^{+i\frac{n\pi}{a}x} - e^{-i\frac{n\pi}{a}x} \right] e^{-i\omega t} \approx \sin\left(\frac{n\pi}{a}x\right) e^{-i\omega t}. \qquad [3.21]$$

These two wave function equations for electrons that conform to $k = k_n = n\frac{\pi}{a}$ each have a corresponding presence probability, that can be written as:

$$\rho^+ = \Psi^+\Psi^{+*} \propto \cos^2\left(\frac{n\pi}{a}x\right) \qquad [3.22]$$

and

$$\rho^- = \Psi^-\Psi^{-*} \propto \sin^2\left(\frac{n\pi}{a}x\right). \qquad [3.23]$$

For the same given value of k_n, Figure 3.5 shows the presence probability densities denoted ρ^+, ρ^- and ρ, which are for electrons with stationary wave Ψ^+, stationary wave Ψ^-, and progressive wave $[\Psi_k(x,t)] = A\ e^{i(kx-\omega t)}$, respectively.

Given its shape, for ρ the progressive wave is constant. When $k = k_n$, this progressive wave can only exist when neglecting the effect of reflection on a lattice with a spatial period equal to a, i.e. with the zero-order approximation $V = V^0 = 0$. Also, when $V \neq V_0$, this type of wave can only exist when $k \neq k_n$.

For example, when $k = \pi/a$, function Ψ^+ has a spatial period given by $\lambda = \left[\frac{2\pi}{k}\right]_{k=\pi/a} = 2a$, whereas ρ^+ takes on the form $|\Psi^+|^2 \propto \cos^2\frac{\pi}{a}x \propto 1 + \cos\frac{2\pi}{a}x$, in other words its period is determined by the term for a period given by a.

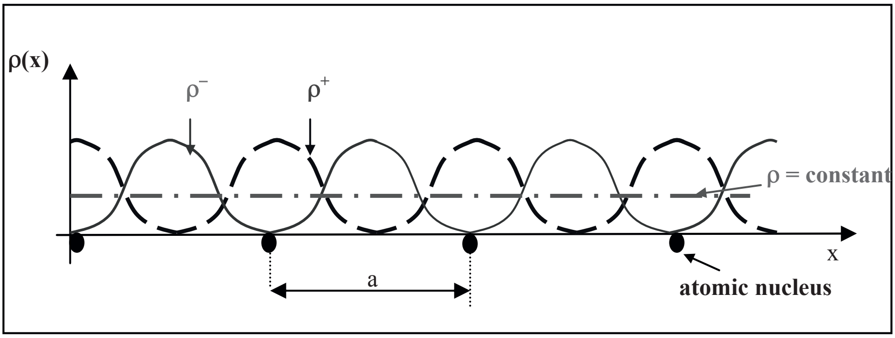


Figure 3.5. *Electron presence probabilities* ρ^+, ρ^-, *and* ρ *for the stationary waves* Ψ^+, Ψ^- *(when* $k=\pi/a$*) and the progressive wave, respectively*

Figure 3.5 shows that when $k = \pi/a$:

– ρ^+ has its maximum concentration of electrons close to the nuclei. This configuration has the lowest average energy (w^+) as the Coulombic energy is negative and has a high modulus due to the short distance between the electrons and nuclei;

– ρ^- has a maximum concentration of electrons midway between nuclei. This configuration has the highest energy (w^-) as the Coulombic potential has a low modulus and the distance between electrons and nuclei is greatest;

– ρ is equal to a constant due to an equal spread of electrons at an intermediate distance from nuclei and, correspondingly, displays an intermediate energy which approximates to that of a free electron.

The existence of two physical solutions, Ψ^+ and Ψ^-, with the same value of k_n ($k = k_n = n\frac{\pi}{a}$) generates two values for energy when in the presence of a periodic potential such as $P(x)$. Figure 3.6 shows the energy dispersion curve, i.e. $E = f(k)$. In the zero-order approximation (free electron with $P(x) \equiv V^0 = 0$), the problem is a degenerate one with solutions in sine and cosine for the same single energy $E^0 = \hbar^2 k^2 / 2m$.

In the presence of a periodic potential, the gap between the two energy values is equal to an energy "gap" given by $E_G = \Delta E^- - \Delta E^+ = w^- - w^+$. This is the so-called forbidden band as there is a flip (for the same value of $k = k_n$) from the energy w^+ to w^-. The calculation of E_G using perturbation theory is detailed in the following section.

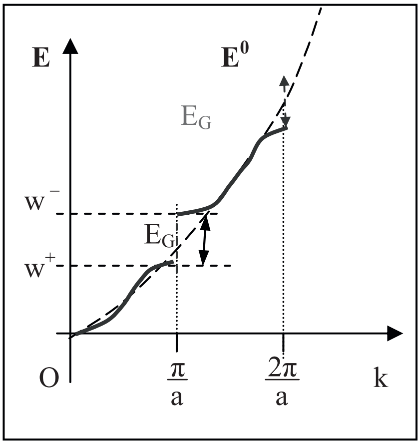


Figure 3.6. *Plot of $E = f(k)$. A zero order approximation (which assumes a perturbation potential of $V(x) = 0$ as interactions between electrons and the lattice are neglected) results in an energy given by $E^0 = \frac{\hbar^2 k^2}{2m}$. The effect of the lattice, through reflection at atoms in the periodic chain of electron waves, is to give two energy solutions ($w^+ < w^-$) for each value of k, as in $k = k_n = n\frac{\pi}{a}$.*

If k is significantly different from the values given by $k_n = n\frac{\pi}{a}$, the wave function takes on a progressive form with two types of propagation being possible (towards the left or the right), each having the same energy value. This is the degenerate value of E^0 when $V = V^0 = 0$, and it remains degenerate in the presence of the periodic potential $V(x) \neq 0$ and has a value given by equation [3.8].

3.4.3. *Results given by the perturbation theory*

3.4.3.1. *Simple estimation of the size of the forbidden band, with $k_1 = \frac{\pi}{a}$*

The functions Ψ^+ and Ψ^-, normalized over a 1D chain containing N atoms such that $N + 1 \approx N$ and thus of length $L = N a$, are such that, with $k_1 = \frac{\pi}{a}$:

$$\int_0^{L=Na} A^2 \left|\Psi^+(x)\right|^2 \mathrm{dx} = 1 \text{, so that } \Psi^+(x) = \sqrt{\frac{2}{L}} \cos\left(\frac{\pi}{\mathrm{a}}x\right)$$

$$\int_0^{L=Na} A^2\left|\Psi^-(x)\right|^2 \mathrm{dx} = 1\text{, so that } \Psi^-(x) = \sqrt{\frac{2}{L}}\sin\left(\frac{\pi}{\mathrm{a}}x\right).$$

The energy gap is thus equal to $E_G = w^- - w^+$, so that on applying the result gained from the perturbation theory presented in section 3.4.1, we obtain:

$$E_G = \Delta E^- - \Delta E^+ = w^- - w^+ = \left\langle \Psi^- \middle| V(x) \middle| \Psi^- \right\rangle - \left\langle \Psi^+ \middle| \mathrm{V(x)} \middle| \Psi^+ \right\rangle$$

$$= \frac{2w_0}{L}\int_0^L \cos\left(\frac{2\pi}{\mathrm{a}}x\right)\left[\sin^2\frac{\pi}{\mathrm{a}}x - \cos^2\frac{\pi}{\mathrm{a}}x\right]dx$$

$$= -\frac{2w_0}{L}\int_0^L \cos^2\left(\frac{2\pi}{\mathrm{a}}x\right)dx = -\frac{2w_0}{L}\int_0^L\left[\frac{1+\cos\left(\frac{4\pi}{a}x\right)}{2}\right]dx = -w_0.$$

Finally, we obtain $E_G = -w_0$. This is a positive value as w_0 is negative, as can be seen in Figure 3.1, where the energy curve reaches a minimum at the nuclei.

Thus, with $E_G = |w_0|$, we can conclude that the higher the value of $|w_0|$, in other words the stronger the electron-lattice interactions, the greater the forbidden band. The origin of the forbidden band is thus from the rise of degeneracy in $k = \frac{\pi}{a}$ from the initial energy level $E^0 = \hbar^2\frac{\frac{\pi^2}{a^2}}{2m}$ caused by periodic potential $V(x)$. The two proper functions $\psi^-(x)$ and $\psi^+(x)$ thus have corresponding energy levels denoted w^- and w^+ separated by $E_G = |w_0|$.

3.4.3.2. *Degeneracy of the restricted problem, with* $k \neq \frac{\pi}{a}$

With these values of k, the wave function is progressive. In the zero order, the wave is given by:

$$\Psi_k^+(x,t) = Ae^{ikx}e^{-i(E/\hbar)t} = \psi_k^+\ e^{-i(E/\hbar)t}$$

or

$$\Psi_k^-(x,t) = A e^{-ikx} e^{-i(E/\hbar)t} = \psi_k^- \, e^{-i(E/\hbar)t},$$

where $A = 1/\sqrt{L}$ for normalized wave functions that satisfy $1 = \left\langle \psi_k^\pm \middle| \psi_k^\pm \right\rangle = \int_0^L |A|^2 \, dx = |A|^2 L$. The energy correction made by inserting the (cosinusoidal) potential of the lattice is zero, so that in effect we have:

$$\Delta E^+ = \left\langle \psi_k^+ \middle| V(x) \middle| \psi_k^+ \right\rangle = \int_0^L |A|^2 e^{ikx} V(x) e^{-ikx} \, dx$$

$$= |A|^2 \int_0^L V(x) dx = |A|^2 w_0 \int_0^L \cos\frac{2\pi}{a} x dx \overset{L=Na}{=} 0$$

and similarly

$$\Delta E^- = \left\langle \psi^- \middle| V(x) \middle| \psi^- \right\rangle = |A|^2 \int_0^L V(x) dx = 0.$$

We thus finally have:

$$\Delta E^+ = \Delta E^- = \Delta E^\pm = 0.$$

There is no longer a rise in degeneracy for k values (in contrast to when $k = \pi/a$) as now with $k \neq \frac{\pi}{a}$ we have $E^+ = E^- = E^0 + \Delta E^\pm = E^0$.

When k is very different from π/a, this result is practically in agreement with that given by equation [3.12] where the second term is in effect negligible.

When k is close to π/a, the calculation becomes too crude to give a precise expression for the energy where the second term of equation [3.12] is non-negligible (small denominator). We must therefore abandon the use of the zero-order wave function in calculations using the perturbation method and turn to equation [3.10] where a first-order wave function is used, even though the use of the expression is in itself debatable for values of k close to π/a. For these values of k, where the use of stationary solutions could be justified, the perturbation method still does not allow a more exact determination of the energy as the stationary solutions are periodic and the energy corrections terms are thus calculated as being zero.

3.4.4. *Conclusion*

To conclude, we can see that the use of perturbation theory is useful when $k = \frac{\pi}{a}$ as it makes a simple calculation of the gap value possible, and enables us to physically determine the origin of the forbidden bands. When $k \neq \frac{\pi}{a}$, the perturbation method used in a simple calculation is not sufficient, and it is rather the resolution of the Mathieu equation that is required.

Put simply, we can note that at the level of zero-order approximation, the solutions for the wave function are given by $\psi = Ae^{\pm ikx}$ (where the ± sign comes from degeneracy of the problem that in turn results from the physical possibility that the electronic wave can propagate in one direction or another), and the stationary solution $\psi_c = \cos(kx)$ or $\psi_s = \sin(kx)$ is obtained by a linear combination of the two exponential solutions. We can also note that these stationary solutions can be seen as real or purely imaginary parts of progressive solutions. In fact, the stationary solutions can only be obtained when the wavelengths (λ_n) of particular electrons in a 1D lattice satisfy the relation $n\lambda_n = 2a$ so that $k_n = n\frac{\pi}{a}$ from equation [3.19]. Given that it is these waves that are reflected from the crystal, the electrons cannot propagate as their wavelengths are commensurable with the lattice period. The result is the establishment of a system of stationary waves (one with sine and the other with cosine forms).

So, when the wavelength λ of an electronic wave is large, or rather, $k = 2\pi/\lambda$ is small, the electrons only weakly sense the effect of the periodic potential $V(x)$ generated by the atom cores (shown in Figure 3.7a). The wave functions (Figure 3.7b) thus evolve relatively free of $V(x)$ and at the limit ($\lambda \to \infty$) can be thought of as being completely disconnected. This results in a constant presence probability across the whole lattice (see also Figure 3.5, which shows how the density $\rho = \text{constant}$ as a solution for a free electron).

In contrast, when λ is small and k moves towards the values $k = \pm\pi/a$, the functions ψ_c and ψ_s (see section 3.4.2) tend towards ψ^+ and ψ^- (with period $2a$ as shown in Figure 3.7c). For these functions, amplitudes of the presence probabilities (of period a) evolve exactly with variations in $V(x)$ as these probabilities are centred on or between atoms (see Figure 3.5). These two very different functions, represented physically as presence probabilities with maximums out of phase by $a/2$ when $k = \pi/a$, corresponding to two different energies that are separated by an "energy gap" denoted by $\Delta E = E_G$ and shown in Figure 3.6.

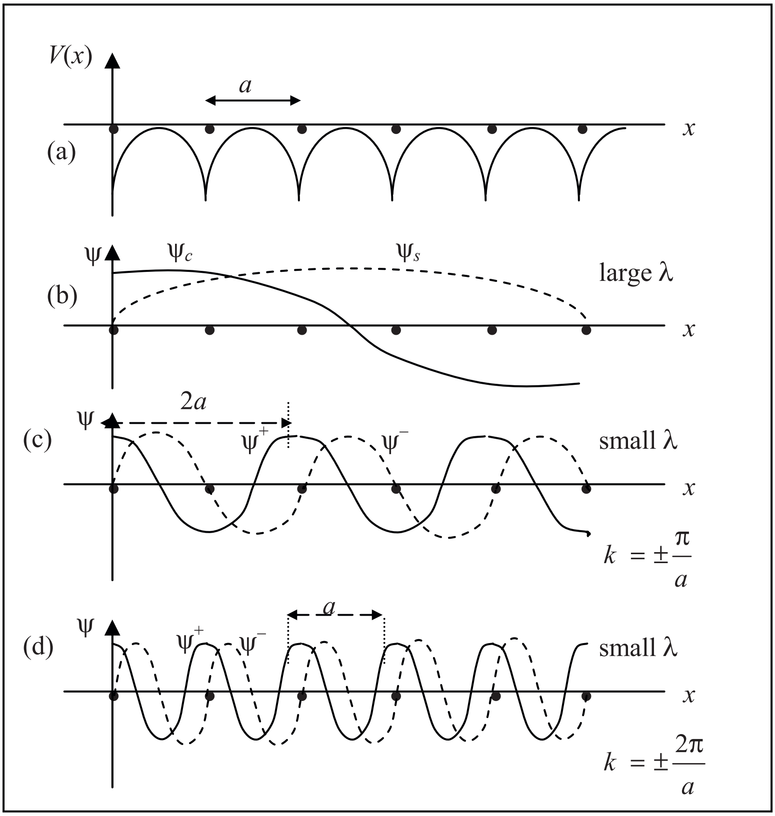


Figure 3.7. *(a) Periodic potential V(x); (b) functions* ψ_c *and* ψ_s *when* λ *is large; (c) functions* ψ^+ *and* ψ^- *when* λ *is small and* $k = \pi/a$*; and (d) functions* ψ^+ *and* ψ^- *when* λ *is small and* $k = 2\pi/a$

When $k = \pm\ 2\pi/a$, functions $\psi^+ \approx \cos\left(\frac{2\pi}{a}x\right)$ and $\psi^- \approx \sin\left(\frac{2\pi}{a}x\right)$ have the period given by a while the period of their presence probability is equal to $a/2$.

In general terms, when $k = \pm n\pi/a$, the period of the wave functions is $2a/n$ ($= 2\pi/|k|$) and that of the presence probability is a/n. The wave functions and, more particularly, their presence probabilities (for the different k values) are in perfect accordance with the terms in $\cos n\omega x = \cos 2\pi n \frac{x}{a}$ (of period a/n) of the Fourier series development for the periodic potential interacting with the lattice $P(x)$ (see section 3.3) and give rise to degeneracy in $k = \pm n\pi/a$.

3.5. Complementary material: the main equation

The central equation, as far as band theory is concerned, is the Schrödinger equation in which the crystalline potential $V(x)$ can, in its most general form, be written as a periodic potential (with the period being that of the lattice) and where $\psi_k(x)$ is a Bloch function. The Schrödinger equation is the most widely applied in band theory.

3.5.1. *Fourier series development for wave function and potential*

In 1D we have $V(x) = V(x + a)$, that is, a periodic function with a period that can be developed as a Fourier series:

$$V(x) = \sum_{m=-\infty}^{+\infty} v_m e^{iG_m x}, \qquad \text{with} \begin{cases} \mathrm{G_m} = \dfrac{2\pi}{a} m \\ v_m = \dfrac{1}{a} \displaystyle\int_{x_0}^{x_0+a} e^{-iG_m x} V(x) dx. \end{cases}$$

Taking into account the symmetry of the problem, as given by $v_m = v_{-m}$, and by choosing an origin for the potentials as being such that $v_0 = 0$, we can write that:

$$V(x) = v_0 + \sum_{\substack{m>0 \\ m=1\to+\infty}} v_m \left(e^{iG_m x} + e^{-iG_m x}\right) = 2\sum_{m=1}^{\infty} v_m \cos G_m x.$$

For its part, wave function $\psi_k(x)$ is a Bloch function that is periodic but where the period is equal to L, i.e. the complete length of the crystal. So, by using the periodic Born von Karmen conditions we have:

$$\psi_k(x) = e^{ikx} u(x) = \psi_k(x + L) = e^{ik(x+L)} u(x + L).$$

As also shown in problem 1 in section 3.6.1, this function $\psi_k(x)$ does not contain the period a. As $u(x)=u(x+a)=u(x+a+a)=\cdots=u(x+Na)$ [with $L=Na$], we have $e^{ikL}=1$, so that $k\equiv k_n=(2\pi/L)n$, and the wave function $\psi_k(x)$ is thus written as:

$$\psi_k(x)=\sum_{n=-\infty}^{+\infty}c_n e^{ik_n x} \qquad \text{with} \begin{cases} \mathrm{k_n}=\dfrac{2\pi}{L}n \quad \text{and n}=0,\ \pm 1,\ \pm 2\ldots \\ \mathrm{c_n}=\dfrac{1}{L}\displaystyle\int_{x_0}^{x_0+L} e^{-ik_n x}\psi_k(x)dx. \end{cases}$$

3.5.2. *Schrödinger equation*

This equation, $-\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2}+V\psi=E\psi$, becomes, on substituting the developments for V and ψ, and on noting that $\frac{\partial^2\psi}{\partial x^2}=-\sum_n k_n^2 c_n e^{ik_n x}$:

$$\frac{\hbar^2}{2m}\sum_n k_n^2 c_n e^{ik_n x}+\sum_m v_m e^{iG_m x}\sum_n c_n e^{ik_n x}$$
$$=\frac{\hbar^2}{2m}\sum_n k_n^2 c_n e^{ik_n x}+\sum_m\sum_n v_m c_n e^{i(k_n+G_m)x}=E\sum_n c_n e^{ik_n x}.$$

In most classic textbooks $k_n=K$ and $G_m=G$, so that the preceding equation is written as:

$$\frac{\hbar^2}{2m}\sum_K K^2 c_{(K)}e^{iKx}+\sum_G\sum_K v_G c_{(K)}e^{i(K+G)x}=E\sum_K c_{(K)}e^{iKx}.$$

Then we make $\frac{\hbar^2}{2m}K^2=\lambda_K$. It should be noted that the preceding equation can only be satisfied if the coefficients in front of the same Fourier component are identical. The coefficient e^{iKx} is obtained for the left member's second term when K also assumes the value $(K-G)$ (because $[e^{i(K+G)x}]_{K\equiv K-G}=e^{iKx}$). For each given value of K, we should therefore satisfy the following equation:

$$\lambda_K c_{(K)}+\sum_G v_G c_{(K-G)}=Ec_{(K)},$$

so that:

$$(\lambda_K - E)c_{(K)} + \sum_G v_G c_{(K-G)} = 0.$$

This is called the central equation due to its pivotal role in solid state physics. Given that the addition over G should, *a priori*, account for an infinite number of terms, its resolution may at first appear difficult. However, the rigid decrease in the coefficient v_G means that only treatment of the first two or four G vectors of the reciprocal lattice are required. If in the development of $V(x)$ we only bring in the two smallest G vectors, given by $G \equiv \pm g = \pm\frac{2\pi}{a}$, the only v_G coefficients which come into play are $v_{\frac{2\pi}{a}} = v_{-\frac{2\pi}{a}} = V$.

3.5.3. *Solution*

For a given value of k in the first zone, the group of equations for the system only account for the coefficients denoted $c_{(K)}$, $c_{(k-Gi)}$ which represent the different vectors of the reciprocal lattice. The initial problem is therefore to resolve N independent problems, each one for a permitted value of k in the first zone.

Finally, the possible values for the energy will be obtained by writing that the determinant of the preceding linear system of equations must be equal to zero. This is because the system of linear equations under consideration (that carry unknowns given by the coefficients denoted by c) only have solutions if the determinant is zero.

3.6. Problems

3.6.1. *Problem 1: a brief justification of the Bloch theorem*

The wave function of a semi-free electron placed within a periodic potential (and of a period such that the dimension of the crystal is given by $L = N\,a$) is given by $\psi_k(x) = e^{ikx}u(x)$, where e^{ikx} is the zero order solution (free electron). To determine the condition that the function $u(x)$ must satisfy:

1) Write the condition that imposes the interaction potential on electron presence probability. From this deduce the wave function that satisfies the relation $\psi(x+a) = C\,\psi(x)$, in which C is determined and is written for N times over.

2) By comparing the expression obtained in the paragraph above with that giving the progressive boundary conditions (PBC), deduce the condition on $u(x)$.

Answers

1) The periodicity of the potential energy involves the periodicity of the electron presence probability, as the electron perceives the same interaction before and after a displacement that is equal to the lattice period. We can thus write that:

$$\left|\psi(x)\right|^2 = \left|\psi(x+a)\right|^2. \qquad [3.24]$$

This means that the wave function must satisfy a relation of the sort given by:

$$\psi(x+a) = e^{i\lambda}\psi(x) \qquad [3.25]$$

which can be rewritten as

$$\psi(x+a) = C\,\psi(x) \qquad [3.26]$$

having made

$$C = e^{i\lambda}. \qquad [3.27]$$

Equation [3.26] N times over gives us:

$$\psi(x+Na) = C^N\,\psi(x). \qquad [3.28]$$

2) In addition, the periodic Born von Karmen conditions (also called the PBC) can be written for a 1D material of length defined by $L = N\,a$, as:

$$\psi(x+Na) = \psi(x), \qquad [3.29]$$

from which identification with equation [3.28] finally yields $C^N = 1$. C thus appears here as an Nth root of the unit so that:

$$C = \exp\left(i\,\frac{2\pi s}{N} \right), \tag{3.30}$$

where $n = 0, 1, 2, \ldots, N - 1$.

The comparison of equation [3.30] with [3.27] shows that λ must therefore of the form $\lambda = \frac{2\pi}{N} n$.

Finally, seeking $\psi(x)$ in the form

$$\psi_k(x) = e^{ikx} u(x), \tag{3.31}$$

is quite legitimate as the zero-order solution is given by $\psi_k^0(x) = e^{ikx}$, and results in finding:

$$\left.\begin{aligned} \psi_k(x+a) &\overset{(6)}{=} e^{ik(x+a)}u(x+a) = e^{ikx}e^{ika}u(x+a) \\ &\overset{(2')}{=} \mathrm{C}\psi_k(x) \overset{(5)}{=} e^{i\frac{2\pi}{N}n}\psi_k(x) \\ &\overset{(7)}{=} e^{i\frac{2\pi}{N}n}\,\mathrm{e}^{\mathrm{ikx}}\,\mathrm{u(x)} \end{aligned}\right\} \Rightarrow \mathrm{u(x)} = u(x+a)$$

with $ka = \frac{2\pi}{N} n$, so that k was of the form $k_n = \frac{2\pi}{Na} n = \frac{2\pi}{L} n$.

3.6.2. *Problem 2: comparison of E(k) curves for free and semi-free electrons in a representation of reduced zones*

The representation of $E(k)$ within a system of smaller zones, with the help of a translation of modulus $n\frac{2\pi}{a}$ (where n is a positive or negative integer), makes its tracing for a domain possible in k where $k \in \left[-\frac{\pi}{a}, +\frac{\pi}{a}\right]$.

1) In question 2 below, where we will see if $\vec{k}\,'$ is a wave vector outside of the interval $\left[-\frac{\pi}{a}, +\frac{\pi}{a}\right]$, the use of a translation of modulus $\left|\vec{A}\right| = n\frac{2\pi}{a}$ makes it possible to bring the extremity of the vector $\vec{k}$ (defined by $\vec{k} = \vec{k}\,' + \vec{A}$) into the interval $\left[-\frac{\pi}{a}, +\frac{\pi}{a}\right]$. To justify such a geometrical transformation, show – in particular – that if in k' the Bloch function is given by $\psi_{k'}(x) = e^{ik'\cdot x} u_{k'}(x)$, and in k if the new wave function $\psi_k(x) = e^{ik\cdot x} u_k(x)$ is still a Bloch function that describes the new state.

NOTE.– In Chapter 5 the chosen vector $\vec{A}$ is in fact a base vector of the reciprocal lattice.

2) Trace the relative positions of the reduced zones for free and semi-free electrons with respect to an initial trace of k in the domain $\left[-3\frac{\pi}{a}, +3\frac{\pi}{a}\right]$.

Answers

1) We first introduce the vector $\vec{k} = \vec{k}\,' + \vec{A}$ in a simplified form for one dimension, $k = k' + A$, into the expression for $\psi_{k'}(x)$ where $u_{k'}(x)$ is a function that has the period denoted a of the lattice and that satisfies the equation $u_{k'}(x) = u_{k'}(x+a)$ as $\psi_{k'}(x)$ is hypothetically a Bloch function.

With $\mathrm{k}' = \mathrm{k} - \mathrm{A}$, we have $\psi_{k'}(x) = e^{ik'x} u_{k'}(x) = e^{ik\cdot x}\left[e^{-iA\cdot x} u_{k'}(x)\right] =$ $e^{ikx} u_k(x)$, if we make $u_k(x) = e^{-iA\cdot x} u_{k'}(x)$.

As $A = n\frac{2\pi}{a}$, we have $e^{-iAx} = e^{-in\frac{2\pi}{a}x} = e^{-in\frac{2\pi}{a}(x+a)} = e^{-iA(x+a)}$, so that $u_k(x) = e^{-iA\cdot x} u_{k'}(x) = u_k(x+a)$. The result is that the function $\psi_k(x) =$ $e^{ikx} u_k(x)$ is a Bloch function, as we have shown that $u_k(x)$ has the periodicity of the lattice.

2)

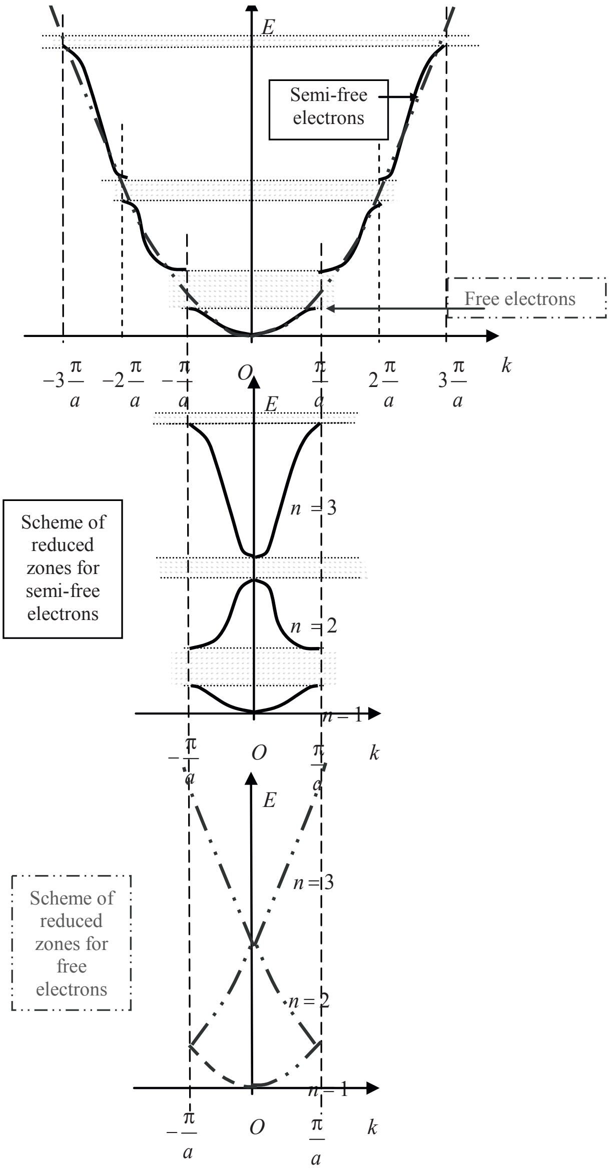


Figure 3.8. *Diagram of reduced zone for free and semi-free electrons*

This page intentionally left blank

Chapter 4

Properties of Semi-Free Electrons, Insulators, Semiconductors, Metals and Superlattices

4.1. Effective mass (m^*)

4.1.1. *Equation for electron movement in a band: crystal momentum*

4.1.1.1. *Preliminary comments: free electrons*

As the internal potential energy of free electrons is zero, the Schrödinger equation is written as $\frac{d^2\psi}{dx^2}+\frac{2m}{\hbar^2}E\psi=0$. We have seen in Chapter 2 that it is possible to write energy (purely kinetic origin with $V=0$) as $E=E_k=\frac{p^2}{2m}=\frac{\hbar^2k^2}{2m}$ (where $\vec{p}=m\vec{v}$) when making $k^2=\frac{2mE}{\hbar^2}$. In fact, we have $k=k_N$ (where N is an integer), and the quantification of k results form the application of the limiting conditions to the function ψ. From the relationship we can deduce that:

$$\vec{p}=m\vec{v}=\hbar\vec{k}. \qquad [4.1]$$

Vector $\hbar\vec{k}$, called the crystal momentum, is equal to the quantity of electron movement when, as here, we are studying electrons. If a free electron is under an external electric field denoted $\vec{E}$, then it is subject to a force given by $\vec{F}_{\text{ext}}=-q\vec{E}$ that results in a variation, given by $d\vec{p}$, in its quantity of movement, itself given by the dynamic fundamental equation. As the total force $(\vec{F}_T)$ experienced by free electrons is only the external force, with no contribution from the internal force, (it is assumed that $V_{\text{interior}} = 0$) then:

$$\vec{F}_{\text{ext}} = \vec{F}_T = m\frac{d\vec{v}}{dt} = \frac{d\vec{p}}{dt} = -q\vec{E}. \qquad [4.2]$$

The introduction of equation [4.1] into [4.2] results in

$$\hbar\frac{d\vec{k}}{dt} = -q\vec{E}. \qquad [4.3]$$

4.1.1.2. *Semi-free electrons*

When these electrons are subject to an external force, we again find $\vec{F}_{\text{ext}} = -q\vec{E}$. The electrons are assumed to move within a permitted band and have a velocity given by $v_g = \frac{d\omega}{dk}$. The Bloch functions being extended over all space means that at a given instant an electron is localized in a wave vector that is not much different from the average wave vector denoted *k*. The permitted energy is given by $E = \hbar\omega$, so that:

$$v_g = \frac{1}{\hbar}\frac{dE}{dk}$$

where *E* is tied to *k* by the so-called dispersion relation $E = E(k)$.

We thus have:

– on one side, $dE = \hbar v_g dk$ (deduced from the expression for v_g);

– and on the other, $dE = -q\mathrm{E}v_g dt$ (obtained after stating that the variation in the energy *dE* is equal to the work produced by an external force given by $-qE$).

From this we have:

$$dk = -\frac{qE}{\hbar}dt,$$

so that in terms of vectors:

$$\hbar \frac{d\vec{k}}{dt} = -q\vec{E} = \vec{F}_{\text{ext}}. \qquad [4.4]$$

Nevertheless, unlike free electrons, the crystal moment for semi-free electrons $\hbar\vec{k}$ differs from the quantity of movement given by $\vec{p} = m\vec{v}_g$. Semi-free electrons are simultaneously subject to both external $(\vec{F}_{\text{ext}} = -q\vec{E})$ and internal $(\vec{F}_{\text{int}})$ forces, the latter being due to the interaction of electrons with internal potential $V(x)$ produced by lattice ions. In place of equation [4.2] for free electrons, the fundamental dynamic equation means that for semi-free electrons we now have:

$$\vec{F}_{\text{ext}} + \vec{F}_{\text{int}} = \vec{F}_T = m\frac{d\vec{v}_g}{dt} \neq \vec{F}_{\text{ext}} \overset{(4)}{=} \hbar\frac{d\vec{k}}{dt}. \qquad [4.5]$$

The upshot is that $\vec{p} = m\vec{v}_g \neq \hbar\vec{k}$ and the quantity of movement of the semi-free electron is different from its crystal moment.

4.1.2. *Expression for effective mass*

With the internal potential and, more importantly, the instantaneous force working on a semi-free electron being unknown for an external observer, we can write an equation that replaces the dynamic fundamental equation. This new equation ties the excitation (the electric field $\vec{E}$ that generates the external force given by $\vec{F}_{\text{ext}} = -q\vec{E}$) to the response of the electron in terms of its displacement (characterized by the velocity $\vec{v}_g$):

$$\vec{F}_{\text{ext}} = m^*\frac{d\vec{v}_g}{dt}. \qquad [4.6]$$

The coefficient denoted m^* thus introduced has the dimension of mass and is called the effective mass (or effective masses).

We thus have:

$$\vec{F}_{\text{ext}} \overset{(4)}{=} \hbar\frac{d\vec{k}}{dt} \overset{(6)}{=} m^*\frac{d\vec{v}_g}{dt},$$

from which we derive:

$$m^* = \frac{\hbar}{\frac{dv_g}{dk}}$$

where $v_g = \frac{1}{\hbar}\frac{dE}{dk}$, so that $\frac{dv_g}{dk} = \frac{1}{\hbar}\frac{d^2E}{dk^2}$, and finally we obtain:

$$m^* = \frac{\hbar^2}{\frac{d^2E}{dk^2}}. \qquad [4.7]$$

NOTE 1.– For a free electron where $E = \frac{\hbar^2 k^2}{2m}$, we have $\frac{d^2E}{dk^2} = \frac{\hbar^2}{m}$, and equation [4.7] gives $m = m^*$ in a result that would seem physically correct, given that for a free electron we have $F_{\text{int}} = 0$, or rather $F_{\text{ext}} = F_{\text{Total}} = m\,\gamma = m^*\gamma$.

NOTE 2.– For a 3D problem, if the function $E(k)$ is anisotropic in space, then the effective mass is a tensor with the components $m^*_{ij} = \frac{\hbar^2}{d^2E/dk_i dk_j}$, where i and j are Cartesian coordinates ($i, j = x, y, z$).

4.1.3. *Sign and variation in the effective mass as a function of k*

4.1.3.1. *Preliminary comment: the curve of E = f (k) for the dispersion relation presents a horizontal tangent at the zone limit (in $k = n\frac{\pi}{a}$)*

In effect, at the zone limit, the solutions for the wave function are in stationary form (see ψ^+ and ψ^- in Chapter 3) and have no propagation. As a consequence, the velocity of the group at the zone limit is zero:

$$[v_g]_{\text{zone limit}} = 0$$

and so with $\frac{dE}{dk} \overset{E=\hbar\omega}{=} \hbar\frac{\mathrm{d}\omega}{\mathrm{dk}} = \hbar v_g$ we have:

$$\left[\frac{dE}{dk}\right]_{\text{zone limit}} = \hbar\left[v_g\right]_{\text{zone limit}} = 0. \qquad [4.8]$$

A longer, more mathematical demonstration is detailed in problem 1 at the end of this chapter.

4.1.3.2. *Sign of the effective mass at the zone limit*

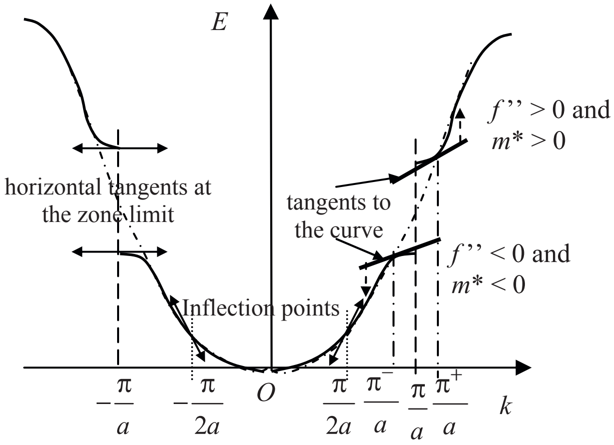


Figure 4.1. *Dispersion curve and sign of the effective mass*

At the zone limit ($k = \pm\frac{\pi}{a}$), the dispersion curves $E = f(k)$ present a horizontal tangent following equation [4.8].

For values of k slightly above $\frac{\pi}{a}$ (in $\frac{\pi^+}{a}$ where we are at the bottom of the band), the curve $E = f(k)$ is above its tangent as at this point the second derivative is positive, i.e. $\frac{d^2E}{dk^2} > 0$, and according to equation [4.7] the effective mass of the electron at the bottom of the band is therefore positive.

For values of k slightly above $\frac{\pi}{a}$ (in $\frac{\pi^-}{a}$ where we are at the top of the band), the curve $E = f(k)$ is below its tangent as at this point the second derivative is negative, i.e. $\frac{d^2E}{dk^2} < 0$, and according to equation [4.7] the effective mass of the electron at the top of the band is therefore negative (Figure 4.1).

4.1.3.3. *Variation in the effective mass as a function of k*

In general, we can state that when the curve $E = f(k)$ gives an inflection point, we have $d^2E/dk^2 = 0$, and for the corresponding values (for example in $k = \pm\pi/2a$ in Figure 4.1), the effective mass tends towards infinity in agreement with equation [4.7]. Equation [4.6] indicates that at this point the electron does not accelerate when acted upon by only the external force $\vec{F}_{\text{ext}}$.

In 3D crystals, m^* only takes on an infinite value in a given direction; in the other directions, its value is finite.

When $k = 0$, the dispersion curves of free and semi-free electrons are the same and $m = m^*$ (Figure 4.2).

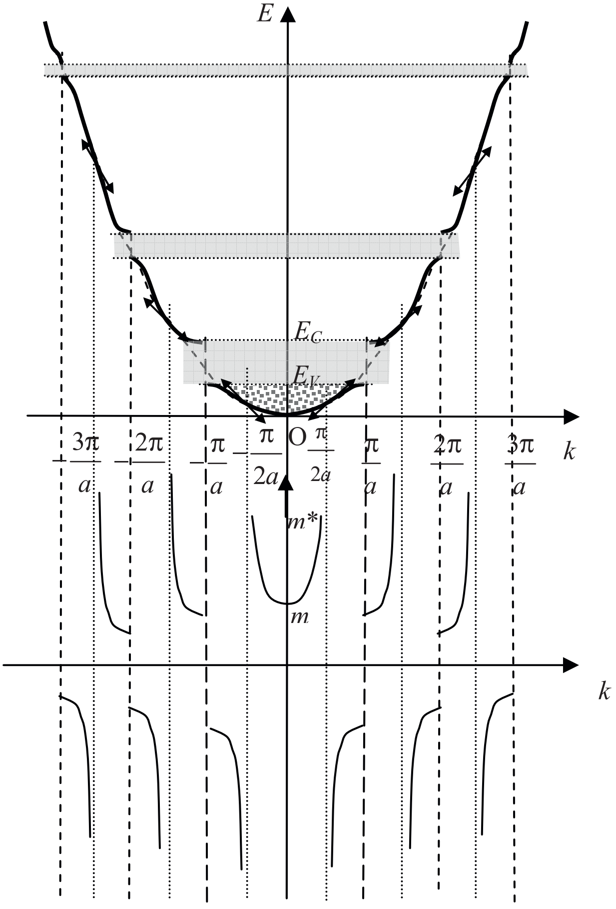


Figure 4.2. *Plot of m* = f(k)*

4.1.4. *Magnitude of effective mass close to a discontinuity*

This subject is also covered in problem 2 later on in this chapter. When effective masses close to a discontinuity are experimentally determined using cyclotron resonance where we can apply:

$$\omega_c = \frac{qB}{m^*}$$

in which B is the magnetic field.

4.2. The concept of holes

4.2.1. *Filling bands and electronic conduction*

The energy levels at absolute zero are progressively filled, bottom up, to the highest permitted energy level, by electrons that are distinguished by their electronic state, that is, by their electronic wave function $\left|\psi_N \ \pm\right\rangle$ determined by their quantum number (N) and spin. If the number of electrons generated by the medium is just sufficient to entirely populate the permitted energy levels of a given band, then this band is described as being full. With all the electrons having been used up, the next band is empty.

In the full band a slight external perturbation, such as an electric field or a thermal excitation that does not carry enough energy to move electrons from the full to the empty band, cannot change the total spread of electrons. This is because there are no energy levels free within the full band to accept electrons. The resultant current density is thus zero. Electrons can swap places but cannot give rise to a resultant electronic transport.

A more mathematical demonstration of this property is shown below. In effect, given the definition of the current density vector we can write that:

$$\vec{j} = -e \sum_{\substack{k \in \text{ full} \\ \text{band}}} \vec{v}(k).$$

Then by using the fact that in 1D we have (see section 4.1.1.2) $v_g = \frac{1}{\hbar}\frac{dE}{dk}$, and vectorially $\vec{v}_g = \frac{1}{\hbar}\overrightarrow{\text{grad}_k}\, E(\vec{k})$, the current density becomes:

$$\vec{j} = -\frac{e}{\hbar} \sum_{\substack{k \in \text{ full} \\ \text{band}}} \overrightarrow{\text{grad}_k} E(\vec{k}).$$

With the functions $E(\vec{k})$ paired in respect of $\vec{k}$ (see Figure 4.1), we have $\sum_{\substack{k \in \text{ full} \\ \text{band}}} \overrightarrow{\text{grad}_k} E(\vec{k}) = 0$, so that the current density produced by a full band is zero, as in $\left[\vec{j}\right]_{\substack{\text{full} \\ \text{band}}} = 0$.

4.2.2. *Definition of a hole*

We will now look at a band that is missing just one electron, that is, a material with a band that when full has n electrons and where the electron is lost from the top of the band, given the filling rules. From the conduction point of view, the state has $(n-1)$ electrons of the material, i.e.:

[material with $(n-1)$ electrons] $=$ [material with n electrons] – [one electron].

The electron removed is from the band summit, and is a particle with charge $-e$ and negative effective mass, as in ($-e$, $-|m^*|$) particle. We can thus write:

[material with $(n-1)$ electrons] = [material with n electrons] – [($-e$, $-|m^*|$) particle].

With the conduction of a material with n electrons being zero (full band), the preceding expression can be rewritten simply in terms of conduction:

[material with $(n-1)$ electrons] =– [($-e$, $-|m^*|$) particle].

Taking part of this expression and combining it with the contribution due to the conduction of a particle of charge $+e$ and with effective mass $+|m^*|$, i.e. ($+e$, $+|m^*|$) particle we can then write that in terms of conduction:

[material with $(n-1)$ electrons] = –[($-e$, $-|m^*|$) particle] – [($+e$, $+|m^*|$) particle] + [($+e$, $+|m^*|$) particle].

The first two terms of the second member can in effect be stated as:

$-[(-e, -|m^*|)$ particle$] - [(+e, +|m^*|)$ particle$] = -[(-e +e, -|m^*| +|m^*|)$ particle$] \equiv 0$, as this particle has zero charge and zero effective mass.

The definitive conduction term can thus be written:

[material with $(n-1)$ electrons] $\equiv$ $[(+e, +|m^*|)$ particle$]$.

It is this particle with a charge denoted $+e$ and an effective mass $+|m^*|$ that is called a hole. The generalization of the preceding reasoning makes it possible to state that if p electrons are removed from the energy states at the top of the band in a material, then we have, in terms of conduction, a material with p holes.

Figure 4.3 shows, for a given electric field going from left to right, the direction of the vectors $\vec{F}_{\text{ext}}, \vec{\gamma}, \vec{v}, \vec{j}$ associated with the transport of electrons and holes (at different energy levels in the permitted bands). We can see straight away that electron current density vectors at the base of the permitted bands and of holes at the top of the band (of course) are going in the same direction.

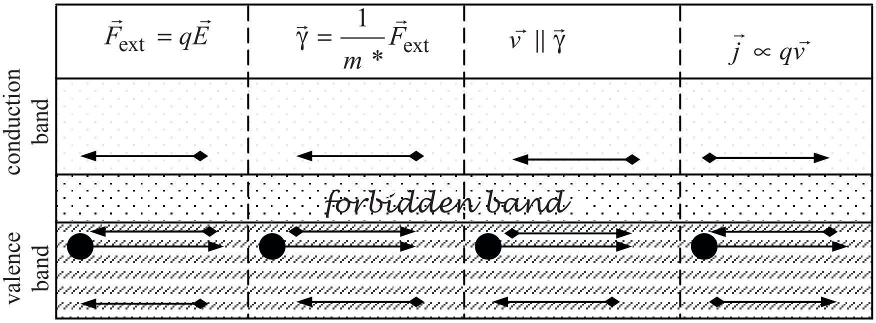


Figure 4.3. *An electric field applied from the left towards the right ($\vec{E}$ ⟶). The direction of the vectors $\vec{F}_{\text{ext}}, \vec{\gamma}, \vec{v}, \vec{j}$ is shown by:*

– the arrow ◆——► for electrons with charge $q = -e$ and effective mass $m^* < 0$ at the band summit and $m^* > 0$ at the bottom of the permitted band; and

– the arrow ●——► for holes of charge $q = +e$ situated by definition at the summit of the permitted band (effective mass $+|m^*|$)

4.3. Expression for energy states close to the band extremum as a function of the effective mass

4.3.1. *Energy at a band limit via the Maclaurin development (in k = $k_n = n\frac{\pi}{a}$)*

Around $k = k_0$ the Maclaurin development for the energy (E) can be written as:

$$E_{(k)} = E_{(k_n)} + (k - k_n)\left(\frac{\partial E}{\partial k}\right)_{k_n} + \frac{(k - k_n)^2}{2}\left(\frac{\partial^2 E}{\partial k^2}\right)_{k_n} . \qquad [4.9]$$

As in $k = k_0$ we have a horizontal tangent (as $\left[\frac{dE}{dk}\right]_{k=k_n} \overset{[4.8]}{=} 0$) while the effective mass is defined by $m^*=$ $m^*=\frac{\hbar^2}{(\partial^2 E/\partial k^2)}$ [4.7], we obtain using equation [4.9] when $k \approx k_n$:

$$E_{(k)} = E_{(\mathrm{k_n})} + \frac{\hbar^2}{2\mathrm{m}^*}(k - k_n)^2. \qquad [4.10]$$

Note that this is the expression that replaces the expression $E = \frac{\hbar^2 k^2}{2m}$ obtained for free electrons.

NOTE 1.– In the neighborhood of the zone center, where n = 0 and $k_0 \approx 0$, we thus find:

$$E_{(k)} = \mathrm{E}_{(\mathrm{k_0})} + \frac{\hbar^2}{2\mathrm{m}^*}k^2. \qquad [4.11]$$

NOTE 2.– Equation [4.10] for the energy of an electron can be given as a function of $|m^*|$. In this case we can write that:

$$E_{(k)} = \mathrm{E}_{(\mathrm{k_n})} \pm \frac{\hbar^2}{2|\mathrm{m}^*|}(k - k_n)^2 \qquad [4.12]$$

where the + sign is for $m^* > 0$, that is, an electron at the bottom of the band, while the – sign is for $m^* < 0$, meaning an electron high up the band.

4.4. Distinguishing insulators, semiconductors, metals and semi-metals

4.4.1. *Required functions*

Insulators, semiconductors and metals are discerned by the way in which the energy levels are filled by their electrons, the latter being characterized with respect to their electronic state (orbital wave function and spin, or so called "spin-orbit" as in $\left|\psi_N \; \pm\right\rangle$). For a rigorous treatment, we should determine the (available) energy state density function for semi-free electrons, denoted $Z(E)$ and the occupation probability function, similarly, denoted $F(E)$ for these energy states in the different bands. The function produced, $n(E) = Z(E)F(E)$, thus represents the density function of occupied (full) electronic states for each energy level (E). Also, $n(e)dE$ represents the number of electrons per unit volume for which the energy is between E and $E + dE$.

In an original and simplifying step, we can state that for $n(E)$ possible filling configurations, each results in a particular electronic behavior and these behaviors can be classed with respect to four types of material, i.e. insulators, semiconductors, metals and semi-metals.

4.4.2. *Dealing with overlapping energy bands*

Generally, in a representation of $E = f(k)$, shown in the upper part of Figure 4.2, the last band to be totally occupied by electrons is called the valence band. Its highest energy level is denoted E_V. The first empty or only partially occupied band is called the conduction band. Its lowest energy level is signified by E_C. However, when looking at 2D or 3D materials, it is no longer possible to see E simply as a function of $k \equiv k_x$ as in Figure 4.2 (that is, traced for a 1D system). In effect, the propagation wave vector $\vec{k}$ for each electron can be specified for each direction of propagation which might be listed as 1, 2, 3 and so on. This means that for the directions x, y and z of a material with optical/electronic axes Ox, Oy and Oz, there are different k values, i.e. k_1, k_2, k_3, etc. The upshot is that different representations of $E = f(k)$ must be made for each direction. For example, say that direction (1) corresponds to the direction x used exclusively until this point. Along with this we have another direction (i) that has no specific alignment. In Figure 4.4 we can see how the representation of $E = f(k)$ can give rise to two different positions for the band limit energy levels denoted E_{Vi} and E_{Ci} for direction (1) with respect to the energy levels E_{V1} and E_{C1} for direction (1). Specifically, in case α (Figure 4.4b) we have $E_{V1} < E_{Vi} < E_{C1}$. Thus there is a resultant forbidden band in the crystal, of a size given by $\Delta E = E_{C1} - E_{Vi} = E_G$. In case β (as in Figure 4.4b), we have $E_{Vi} > E_{C1}$ and the forbidden band disappears because the energy levels of the conduction band of direction (1) start filling prior to the valence band of direction (i) being totally filled.

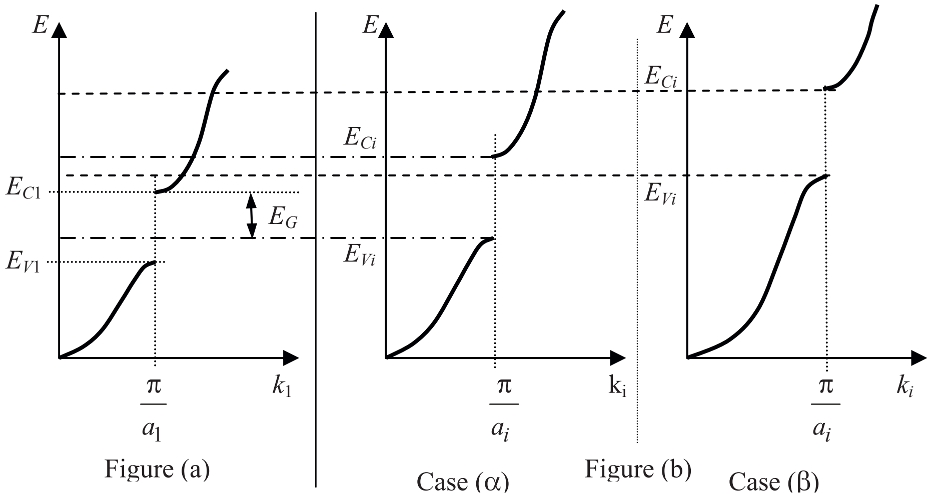


Figure 4.4. *Plots of $E = f(k_i)$ for: (a) values of E_{V1} and E_{C1} fixed by a given direction k1; (b) the resulting cases α and β with different positions of E_{Vi} and E_{Ci}*

Generally, when the forbidden band appears, it has a size given by the minimum value of ΔE obtained from the difference between the minimum value of all the energies at the bottom of the conduction band (E_{Cj}) (for all directions $j = 1, I$, etc. under consideration) and the maximum value reached by the highest energy of the valence band (E_{Vj}) (again for all directions $j = 1, I$, etc.).

4.4.3. *Permitted band populations*

4.4.3.1. *State density function Z(E)*

As we saw for free electrons in Chapter 2, the state density function can be defined either in wavenumber (k) space (reciprocal space) or in the energy space given by E. For the latter and in 3D space, the relevant expression is given by (see also equation [2.31] in Chapter 2):

$$Z(E) = \frac{4\pi}{h^3}(2m)^{3/2}\sqrt{E}. \qquad [4.13]$$

For these free electrons, where there is no interaction potential with the lattice, the function $Z(E)$ is thus parabolic.

For semi-free electrons or holes, the interaction of carriers with the lattice means replacing the mass (m) of free electrons with the effective mass of the carrier under consideration. Semi-free electrons, which have an effective mass denoted by m_e*, are distributed in the conduction band from the bottom up, the bottom being level E_C rather than zero energy, as would be the case for free electrons. The state density function can be calculated just as for free electrons, with the exceptions that m and $(E - 0) = E$ are respectively replaced by m_e* and $(E - E_C)$. Equation [4.13] now takes on a form shown below as equation [4.14]. This equation also shows a parabolic evolution for energies such that $E > E_C$ (see Figure 4.5).

$$Z_C(E) = \frac{4\pi}{h^3}(2m_e^*)^{3/2}\sqrt{(E - E_C)}. \qquad [4.14]$$

Similarly, semi-free electrons are distributed in the valence band from the summit of the band, i.e. E_V downwards so that the state density function is parabolic towards energies, such that $E < E_V$. With function $Z(E)$ still being positive, it is the absolute value of the effective mass that must be brought into play, so that we now obtain:

$$Z_V(E) = \frac{4\pi}{h^3}(2\left|m_e^*\right|)^{3/2}\sqrt{(E_V - E)}. \qquad [4.15]$$

To take these differences into account, if we denote the band limit by E_n (so that E_n is equal to either E_C or E_V), and make m* the general term for the effective mass of the carrier (be it electron or hole), then the general expression for the state density function for all types of carriers is:

$$Z(E) = \frac{4\pi}{h^3}(2|m^*|)^{3/2}\sqrt{|E - E_n|}. \qquad [4.16]$$

We can now represent the overall evolution of the state density function $Z(E)$ for electrons in Figure 4.5, where just as in section 4.4.2 we can distinguish two specific situations:

– case α where there is no band overlapping and there is an energy gap such that $\Delta E = E_G$;

– case β where there is an overlap of valence and conduction bands and as a consequence a suppression of the forbidden energy zone.

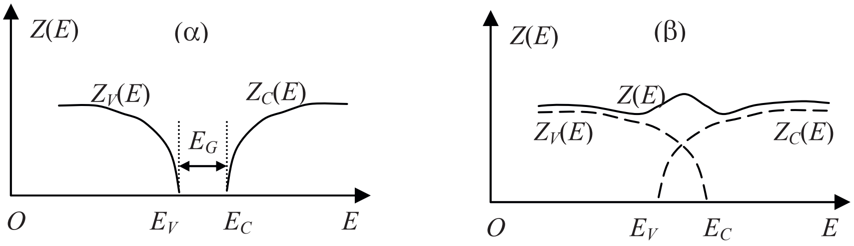


Figure 4.5. *State density function for semi-free carriers where there: (α) is no band overlap; and (β) is band overlap*

4.4.3.2. *Occupation probability function F(E)*

Fermi-Dirac distribution is necessary when describing the distribution of electrons (or holes), as it governs the quantum behavior of a range of specific particles:

- under thermodynamic equilibrium;
- without mutual interactions; and
- likely subjection to an external field.

This is in effect a Boltzmann distribution modified by a condition due to the Pauli principle. Using Fermi statistics (details of which can be found in most courses on thermodynamic statistics) the occupation probability of a given energy level E by electrons is described by the Fermi-Dirac function (where E_F is Fermi energy):

$$F_n(E) = \frac{1}{1 + \exp\left(\frac{E - E_F}{kT}\right)}. \qquad [4.17]$$

This function is such that $F_n(E_F) = 1/2$. When $T = 0K$ we have $F_n\ (E < E_F) = 1$ and $F_n\ (E > E_F) = 0$ (see Figure 4.6b). When $T \neq 0K$, we obtain the result plotted in a continuous line in Figure 4.7b.

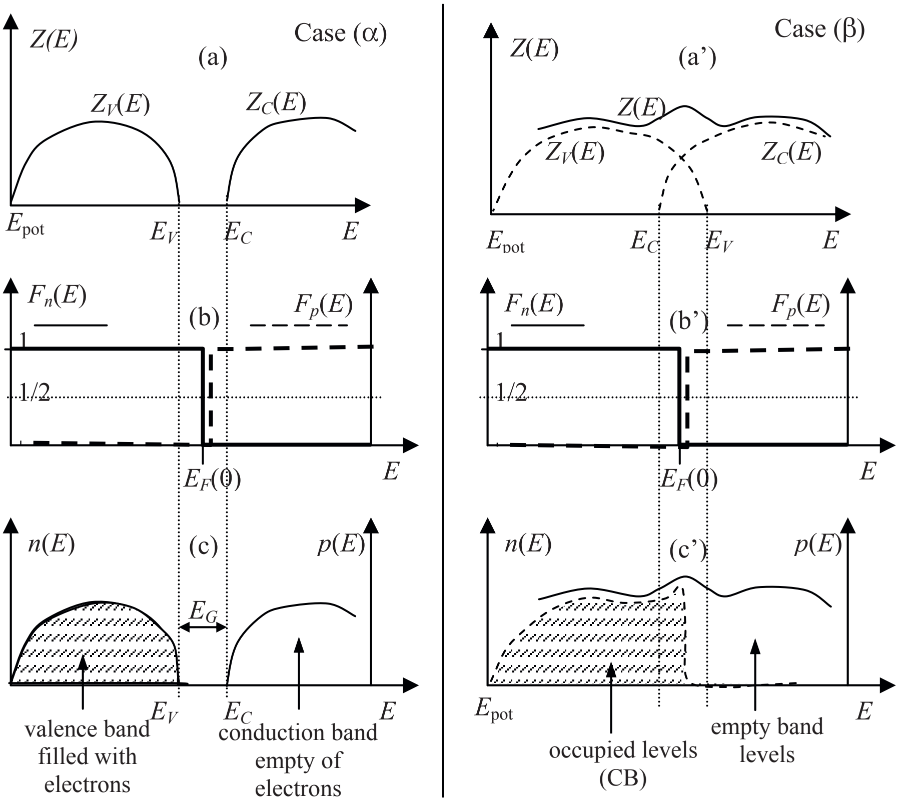


Figure 4.6. *At absolute zero, a qualitative evaluation of the function n(E) from functions Z(E) and F(E) when E_F is localized mainly in the middle of the forbidden band in a system of: (α) non-overlapping bands; and (β) overlapping bands*

To obtain the distribution law for holes denoted $F_p(E)$, we first note that a given energy level E is either occupied by an electron or occupied by a hole (electron vacancy). This trivial condition is transcribed as $F_n(E) + F_p(E) = 100\% = 1$, so that we can state that:

$$F_p(E) = 1 - F_n(E) = 1 - \left[1 + \exp\left(\frac{E - E_F}{kT}\right)\right]^{-1}. \qquad [4.18]$$

This function is such that $F_p(E_F) = 1/2$, and when $T = 0K$ we have $F_p\,(E < E_F) = 0$ and $F_n\,(E > E_F) = 1$ (see Figure 4.6b). When $T \neq 0K$, we find the representation given as a broken line in Figure 4.7b.

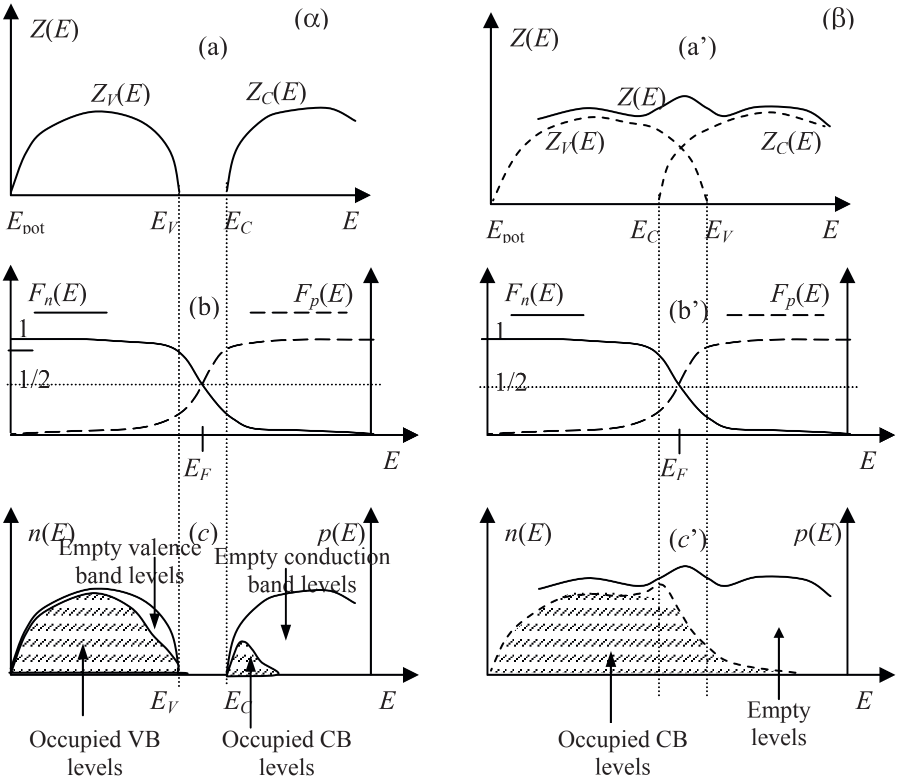


Figure 4.7. *For $T \neq 0K$, qualitative evaluations of the function $n(E)$ from functions $Z(E)$ and $F(E)$ when E_F is well localized in the middle of the forbidden band for a system of bands that: (α) do not overlap; (β) overlap*

4.4.3.3. *Electronic densities in the permitted bands: distinction between insulators, semi-conductors, metals and semi-metals*

If $n(E)dE$ represents the number of electrons per unit volume having energy between E and $E + dE$, the electronic density $n(E)$ is equal to the product of the number of available states between E and $E + dE$ (or rather $Z(E)dE$) and the occupation probability $F_n(E)$ of these states. This can be stated as $n(E) = Z(E)F_n(E)$.

Similarly, for holes we will have a density $p(E)$ given by $p(E) = Z(E)F_p(E)$. The equations for these densities will be developed in the second volume entitled *Electronics and Optoelectronics of Materials and Devices*. For now, we will limit ourselves to qualitatively studying the filling of permitted bands.

4.4.3.3.1. When at absolute zero

In Figure 4.6 the Fermi level is placed broadly in the middle of the gap – a position that will be justified later on for semiconductors – however, it can be found in the permitted band as in the case of certain metals.

Qualitatively, at absolute zero, when there is no band overlap, as in case α on the left, the last occupied band is totally filled (valence band) while the following band is totally empty (conduction band) as in Figure 4.6c. The electronic gap E_G (width of the forbidden band) is thus determined by the difference in energy between the base of the conduction band (E_C) and highest point of the valence band (E_V). When this gap $E_G = E_C - E_V$ is:

– sufficiently large (typically > 5 eV) to stop significant passing of electrons between bands by thermal agitation kT (Figure 4.7), the material is considered to be an insulator;

– sufficiently narrow (typically < 3 eV) to allow significant passing of electrons from the valence to the conduction band by thermal agitation kT (Figure 4.7), the material is considered an *intrinsic semiconductor*.

Qualitatively, at absolute zero, when there is a band overlap, as in case β on the right-hand side of the figure, all the levels below E_F are filled while those above remain empty. This partially filled band is in reality a conduction band. The position of the Fermi level at absolute zero $E_F(0)$ is thus defined by the equality:

$$\int_{E_{\text{pot}}}^{E_F(0)} Z(E)dE = \mathbb{Z} \text{ density of electrons in the conduction band}$$

where E_{pot} is the potential energy at the base of the conduction band. In fact two situations can be distinguished in this latter situation:

– either the overlap of the bands is weak and few states are involved so as to give a *semi-metal*; or

– the overlap of bands is strong with many states involved and we have a *metal*.

Outside the case presented in Figure 4.6 where the Fermi level is in the center of the forbidden band, we should also consider the case where E_F is found within a permitted band. In this situation, the states of the permitted band are filled up to the level E_F and then bands above E_F are empty. With the occupied states not separated from the empty states by a forbidden band, we have a *metal*.

4.4.3.3.2. When $T \neq 0K$

Here the function $Z(E)$ is not modified, while the Fermi function does not suddenly pass from a value equal to one to being equal to zero (and *vice versa*) in $E = E_F$. The change occurs progressively, more so as T increases.

In case α on the left-hand side of Figure 4.7, there is no overlapping of the bands. We can now see that there are states at the top of the valence band (VB) that are unoccupied (partially filled band) while there are some states at the bottom of the conduction band (CB) that are becoming filled. The movement of electrons from the valence to the conduction band is even greater when E_G is small (as in semiconductors) and when the temperature is high. With the total number of electrons remaining constant in a given volume, the number of empty levels in the VB must be equal to the number of filled levels in the CB. Because function $Z(E)$ varies as a parabola in both the heights of the VB as well as in the depths of the CB and function $F(E)$ is symmetric around the Fermi level (at point E_F, ½), the Fermi level must be placed well near the middle of the forbidden band so that the empty and hatched surfaces shown in Figure 4.7c of the VB and the CB, respectively, are equal, so that:

$$E_F \simeq \frac{1}{2}\left(E_V + E_C\right).$$

In case β on the right-hand side of Figure 4.7, where there are overlapping bands, the abrupt transition at $T = 0K$ between the full and empty level is considerably smoothed at $T \neq 0K$. The permitted band levels remain partially empty and the metallic or semi-metallic behavior is retained when $T \neq 0K$, regardless of temperature.

4.4.3.3.3. Conclusion

A normal representation of the band scheme is given in Figure 4.8, where we are now dealing with an intrinsic semiconductor. The electronic and hole state density functions, $n(E)$ and $p(E)$ respectively, are detailed.

Other band representation schemes are used, in particular that of reduced zones (see also section 3.6.2). Instead of representing the dispersion curve $E = f(k)$ over all variations in k (which would bring in various values of k as in $k = \frac{\pi}{a}, 2\frac{\pi}{a}, 3\frac{\pi}{a}$, etc., for which we would see the discontinuities shown in Figure 4.2 for insulators and semiconductors), the $E = f(k)$ is represented simply for values of k between $-\frac{\pi}{a}$ and $+\frac{\pi}{a}$. However, this is also for the whole energy range brought into the zone given by $\left[-\frac{\pi}{a},+\frac{\pi}{a}\right]$. The curves represent energy with the help of an appropriate translational modulation of the form $\pm\, n\,\frac{2\pi}{a}$ (translation modulation equal to that of an appropriate vector of the reciprocal lattice so that the translation conserves its Bloch function properties in each zone. This is further described in Figure 4.9 where the number of each zone is also shown.

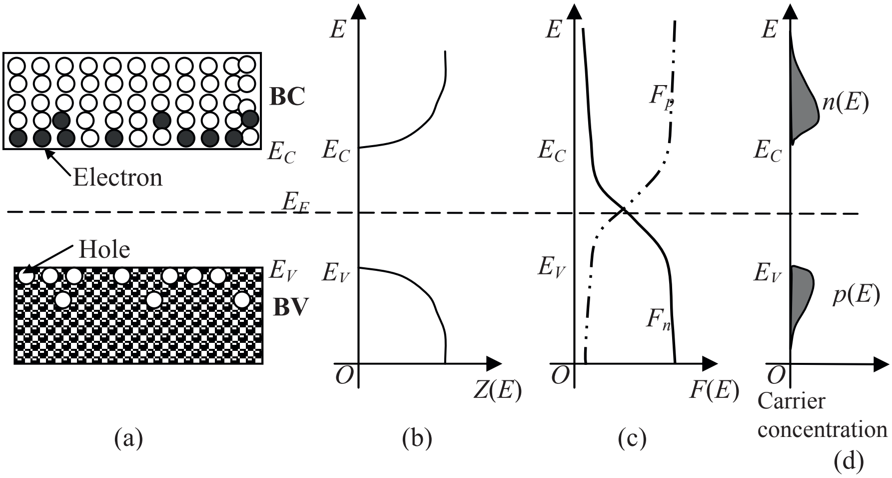


Figure 4.8. *(a) Band filling for a intrinsic semiconductor; (b) state density functions; (c) occupation probability functions; and (d) concentration functions*

The representation shown in Figure 4.9 is that of a 1D medium (see section 3.6.2) when the extremities of the curves are at $k = 0$. In a 2D or 3D media, where the overlapping zone is not necessarily in the same direction, we can end up with the schemes shown in Figure 4.10. In the direction k_x where the lattice repeat unit is denoted by a, the various zones are brought back into the zone defined by the interval $\left[0,+\frac{\pi}{a}\right]$. Similarly, in the k_y direction where the lattice repeat unit is denoted by b, the various zones are brought back into interval $\left[0,+\frac{\pi}{b}\right]$. Figure 4.10a thus represents an insulator/semiconductor (lower band full and upper band empty at $T = OK$), Figure 4.10b for a metal with the last occupied band being partially full (leaving free the upper part of the band where electrons can move when influenced for example by an electric field), and Figure 4.10c for a semi-metal due to the weak overlap of bands.

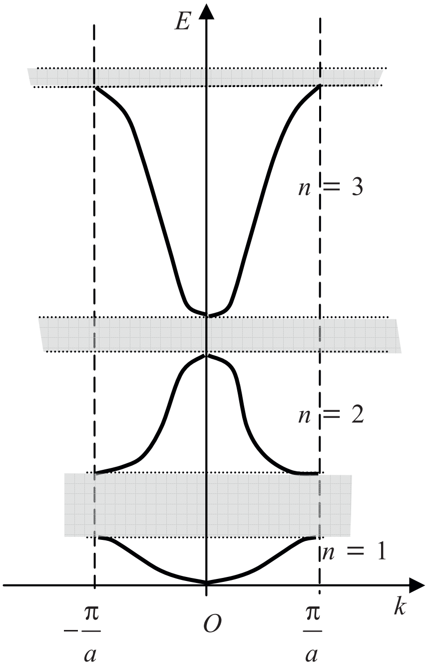


Figure 4.9. *Diagram of reduced zones, where each zone is given an n number*

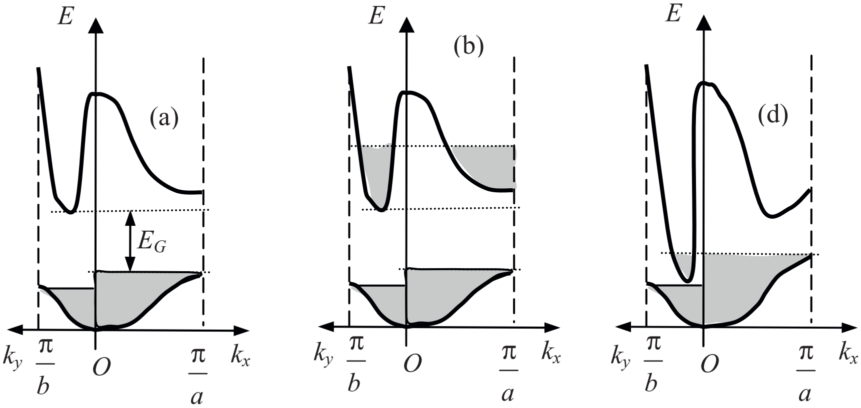


Figure 4.10. *Band structure for: (a) an insulator with a fully occupied last band; (b) a metal where the last occupied band is only partially filled; and (c) a semi-metal with band overlap*

4.5. Semi-free electrons in the particular case of super lattices

A practical example can be given using the problem below.

A heterostructure is based on three pieces of semiconductor (SC) as shown in the figure just below. A thin SC_1 slice of thickness denoted d along Tz and in the order of 10 nm is sandwiched between two layers of the same SC_2. In directions Tx and Ty the L dimensions of SC_1 and SC_2 are considerably greater than d. Finally, the "gaps" of SC_1 and SC_2, respectively denoted as E_{G_1} and E_{G_2}, are such that $E_{G_1} < E_{G_2}$.

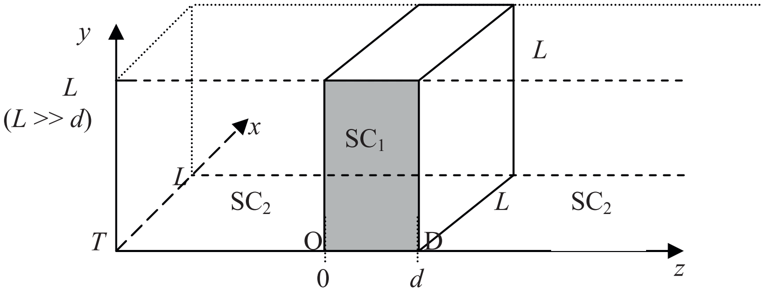


Figure 4.11. *Heterostructure based on three pieces of semiconductor*

The heterostructure can be represented by a theoretical model in which movement is separated into a particular plane of the structure (plane cTy) and in the perpendicular direction (Tz). The function is thus given as:

$$\psi(x, y, z) = \xi(z)\, \Phi(x, y),$$

and it is assumed that the electronic states at the SC_1 level of the SC are those of an electronic gas (at zero potential) in 3D, with an effective mass m^* (which thus takes interactions between electrons and the lattice into account as $m^* \neq m$ where m is the mass of a free electron).

It is also assumed that these electrons:

– move freely in planes parallel to xTy; and

– exhibit limited displacements within zone $0 \leq z \leq d$ due to the existence of potential barriers (that are assumed to be infinite within $z = 0$ and $z = d$ in this theoretical study).

Before approaching the example problems, we should note that if the wave function is the product of two functions, as in $\psi(x, y, z) = \xi(z)\,\Phi(x, y)$, then the energy solution from the Schrödinger equation is given by $E = E_z + E_{x,y}$ and the behavior in this system is the same as that in a classical potential box.

1)

a) Write the Schrödinger equation for the displacement of electrons (in SC_1) in the perpendicular direction (Tz).

b) Give the form of the wave function [$\xi(z)$] using fixed boundary conditions (FBC) that gives a stationary solution and for which the wave function takes nodes as extremities. Justify this choice.

c) Deduce the conditions for quantification (with the help of a quantum number denoted n_z for which the values will be given) of the wave vector (component k_z) and corresponding energy E_z.

d) With m^* (AsGa) = 0.066 m (where m is the ususal mass of the electron) and $d = 10$ nm, give the numerical value of the first three permitted energies E_{z_1}, E_{z_2}, E_{z_3} in eV.

2)

a) Write the Schrödinger equation for the displacement of electrons (in SC_1) in the plane xTy.

b) Give the form of the wave function $\Phi(x, y)$. Justify the use of progressive boundary conditions (PBC, otherwise called Born Von Karman boundary conditions) to determine the quantification conditions (with the numbers n_x and n_y to be determined) of the wave vector (components k_x and k_y). As indicated in the figure, extensions of the semiconductor in directions Tx and Ty are equal to L, where L is considerably greater than d by several orders of size.

c) Deduce the energy values $E_{x,y}$.

3) Give the final expression for the quantified energy of the electrons of SC_1 and determine the minimum value ($E1$) of this energy.

4) Making $k_{//} = \sqrt{k_x^2 + k_y^2}$ and $k_\perp = k_z$:

a) give on the same plot the dispersion curve for $E_{x,y} = f(k_{//})$ and $E_z = f(k_\perp)$, and then $E = f(k_{//})$ and $E = f(k_\perp)$;

b) detail the progressive filling of the permitted level. Determine, in particular, the conditions on n_x and n_y when going from $n_z = 1$ to $n_z = 2$;

5) For direction *Tz*, we suppose that the touching semiconductors have the same reference potential (vacuum level) and it is with respect to this that we place the bottom of the conduction bands (due to electronic affinities χ_1 for SC_1 and χ_2 for SC_2). The structure used in this example is: AlAs/GaAs/AlAs where in eV we have:

E_{G1} (GaAs) = 1.43,
E_{G2} (AlAs) = 2.16,
χ_1 (GaAs) = 4.07,
χ_2 (AlAs) = 3.5.

a) Describe the positions of the conduction band minimums and valence band maximums for structure $SC_2/SC_1/SC_2$ with respect to *Tz*. It should be noted that, if required, the electronic affinity is the difference in energy between the vacuum level and base of the conduction band.

b) Indicate the height of the potential barrier at the interface SC_2/SC_1. Give an approximate position of the above evaluated energy levels E_{z_1}, E_{z_2}, E_{z_3} on the diagram. Detail the nature of approximations made on this result in order to give the approximate energy levels.

c) Also give the position $E_{z1}(t)$ of the holes in GaAs by taking for their effective mass the value $|m^*(t)| = 0.68$ m.

d) From this deduce the radiation energy that might be emitted from GaAs.

Answers

In order to study the electronic states in a superlattice structure, where there are two semiconductors that are such that $E_{G1} < E_{G2}$, we can denote two types of movement:

– one along *Oz* where SC_1 presents a small dimension *d*, for which we will privilege the stationary solutions for the wave function through the use of FBC; and

– one along directions *Ox* and *Oy* where SC_1 has a large dimension with respect to *d* (plan0065 *xTy*). Progressive solutions are favored for the wave function through the use of PBC.

We make $\psi(x,y,z) = \xi(z)\Phi(x,y)$ (separation of variable), and the energy will be in the form $E = E_z + E_{x,y}$.

1)

a) With respect to *Tz*, the Schrödinger equation is written as: $\frac{\partial^2 \xi}{\partial z^2} + \frac{2m^*}{\hbar^2} E_z \xi = 0$.

b) By making $k_z{}^2 = \frac{2m^*}{\hbar^2} E_z$, the stationary solutions for the wave function can be written as: $\xi(z) = Ae^{ikz} + Be^{-ikz}$.

The use of FBC for the direction Oz where SC_1 has a small dimension makes it possible to state that the wave function presents a node at the origin: $\xi(0) = 0$, from which $A + B = 0$, or rather $A = -B$ and from which it can be deduced that $\xi(z) = 2A \frac{e^{ikz} + e^{-ikz}}{2}$, which can also be written in the form: $\xi(z) = 2A \frac{e^{ikz} + e^{-ikz}}{2} = 2iA \sin k_z z$.

Assuming that the wave function has a node at each extremity is the same as assuming that at these extremities the barrier between adjacent semiconductors is practically infinite, which will require verification *a posteriori*.

c) The alternate limiting conditions thus give $\xi(d) = 0$, which results in sin $k_z d = 0$, from which $k_z d = n_z \pi$, so that also $k_z = n_z \frac{\pi}{d}$, where $n_z \geq 1$ and an integer and not equal to zero. The latter is because otherwise the solution would give $\xi(z) = 0$ which would correspond to the absence of particles with a presence probability equal to zero.

We can go on to deduce that $E_z = \frac{\hbar^2 k_z{}^2}{2m^*} = \frac{\hbar^2}{2m^*} \frac{\pi^2}{d^2} n_z{}^2 = \frac{h^2}{8m^* d^2} n_z{}^2$.

d) We find:

$E_{z1} = 0.057$ eV,
$E_{z2} = 0.23$ eV,
$E_{z3} = 0.52$ eV.

2)

a) Schrödinger's equation for the plane xTy is written as:

$$\frac{\partial^2 \Phi(x,y)}{\partial x^2} + \frac{\partial^2 \Phi(x,y)}{\partial y^2} + \frac{2m^*}{\hbar^2} E_{x,y} \Phi(x,y) = 0$$

b) The separation of variable is obtained by making:

$\Phi(x,y) = X(x)Y(y)$,

from which we find energy in the form $E_{x,y} = E_x + E_y$. Placing this in to the Schrödinger equation we obtain the sum for the two equations independently in *x* and *y* directions:

$$\begin{cases} \dfrac{\partial^2 X(x)}{\partial x^2} + \dfrac{2m^*}{\hbar^2} E_x X(x) = 0 \\ \dfrac{\partial^2 Y(x)}{\partial y^2} + \dfrac{2m^*}{\hbar^2} E_y Y(y) = 0. \end{cases}$$

By making $k_x{}^2 = \frac{2m^*}{\hbar^2} E_x$ and $k_y{}^2 = \frac{2m^*}{\hbar^2} E_y$, the search for progressive solutions with respect to these large dimensions *x* and *y* leads us to (with respect to *x*):

$$X(x) = Ce^{ik_x x}.$$

The use of the corresponding PBC for progressive solutions makes it possible to write that: $Ce^{ik_x x} = X(x) \overset{CLP}{=} X(x+L) = Ce^{ik_x (x+L)}$, from which is deduced that $e^{ik_x L} = 1$.

The upshot of this is that we should have $k_x L = 2\pi n_x$ with $n_x \in \mathbb{Z}$. n_x is an integer that can be positive, negative or zero. The last value will simply yield a constant probability presence.

Similarly, for *Oy* we obtain: $k_y L = 2\pi n_y$ with $n_y \in \mathbb{Z}^*$.

c) From the preceding values of k_x and k_y we can deduce the value of the energy for movement in the plane *xTy*:

$$\begin{aligned} E_{x,y} = E_x + E_y &= \frac{\hbar^2}{2m^*}\left(k_x^2 + k_y^2\right) \\ &= \frac{\hbar^2}{2m^*}\left(\frac{2\pi}{L}\right)^2 \left(n_x{}^2 + n_y{}^2\right) \\ &= \frac{h^2}{2m^* L^2}\left(n_x{}^2 + n_y{}^2\right) \end{aligned}$$

3) The final expression for the quantified energy E of electrons in SC$_1$ can thus be written as:

$$\begin{aligned} E &= E_z + E_{x,y} \\ &= \frac{\hbar^2\pi^2}{2m^*}\left[\frac{n_z^2}{d^2} + \frac{4}{L^2}\left(n_x^2 + n_y^2\right)\right] \\ &= \frac{h^2}{8m^*}\left(\frac{n_z^2}{d^2} + \frac{4}{L^2}\left(n_x^2 + n_y^2\right)\right). \end{aligned}$$

The minimum value in E, denoted E_1, for the lowest values in n_z, n_x and n_y, i.e. $n_z = 1$ and $n_x = n_y = 0$ gives:

$$E_1 = (E)_{\substack{n_z=1 \\ n_x=n_y=0}} = \frac{h^2}{8m^* d^2}.$$

4)

a) By making $k_{//} = \sqrt{k_x{}^2 + k_y{}^2}$ and $k_\perp = k_z$, we can trace the dispersion curves for $E_z = f(k_\perp)$ and $E_{x,y} = f(k_{//})$, from which we can deduce the curves $E = E_z + E_{x,y} = f(k_{//})$ and $E = f(k_\perp)$. See the following figure obtained by observing the comments below.

In particular, we can note that:

– on the other hand: $E_z = \frac{\hbar^2}{2m^*}k_z{}^2 = \frac{\hbar^2}{2m^*}k_\perp{}^2 = \frac{\hbar^2}{2m^*}\frac{\pi^2}{d^2}n_z{}^2$, where $n_Z = 1, 2, 3, \ldots$ and $\frac{\pi}{d}$ is large as d is small (with respect to L), so that E_z is large (compared to $E_{x,y}$), and that $(E_z)_{n_z=1} = E_1$;

– on the other hand: $E_{x,y} = \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2) = \frac{\hbar^2}{2m^*}k_{\|}^2 = \frac{\hbar^2}{2m^*}\left(\frac{2\pi}{L}\right)^2(n_x^2 + n_y^2)$ where n_x and $n_y \in \mathbb{Z}$ and $\frac{2\pi}{L}$ is small, as L is large (with respect to d).

The result is that $E_{x,y}$ is small (with respect to E_z), and that $k_{\|}$ can be written as $k_{\|} = 0, \frac{2\pi}{L}, \frac{2\pi}{L}\sqrt{2}, \frac{2\pi}{L}2 = \frac{4\pi}{L}, \ldots$

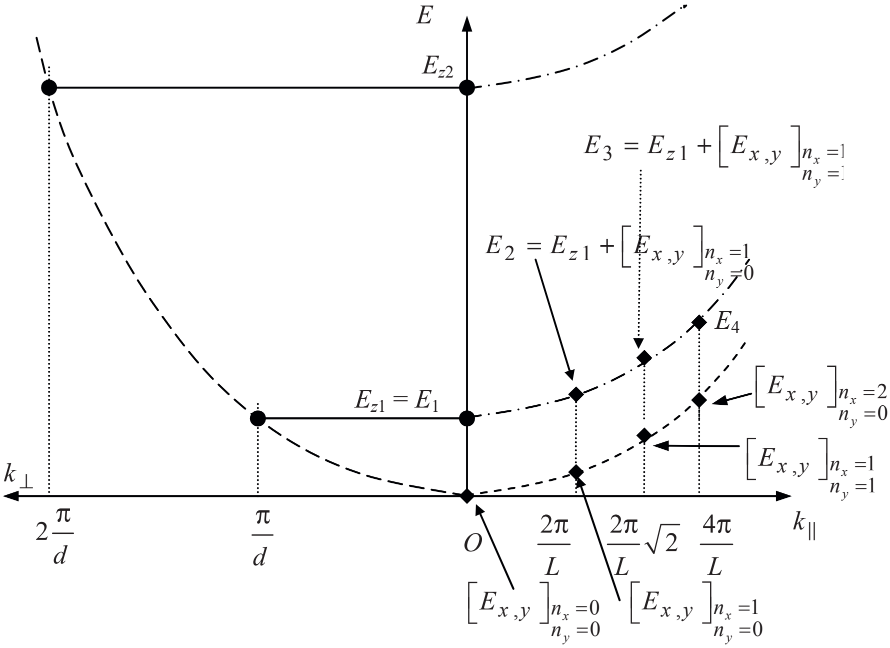


Figure 4.12. *Dispersion curves of* $E_z = f(k_\perp)$ *and* $E_{x,y} = f(k_{//})$ *from which we can obtain the curves of* $E = E_z + E_{x,y} = f(k_{//})$ *and* $E = f(k_\perp)$

b) The progressive occupation of levels is in the order E_1, E_2, E_3, E_4 and so on, as shown in Figure 4.11. As long as $E_{z2} \geq E_{z1} + [E_{x,y}]_{n_x,n_y}$, the obtained energy E remains that obtained with $n_z = 1$, so that:

$$\frac{\hbar^2}{2m^*}\frac{\pi^2}{d^2}(2)^2 \geq \frac{\hbar^2}{2m^*}\frac{\pi^2}{d^2}(1)^2 + \frac{\hbar^2}{2m^*}\left(\frac{2\pi}{L}\right)^2\left(n_x^2 + n_y^2\right),$$

from which we can deduce:

$$3\frac{\pi^2}{d^2} \geq \frac{4\pi^2}{L^2}\left(n_x^2 + n_y^2\right),$$

thus giving the condition: $n_x^2 + n_y^2 \leq \frac{3}{4}\frac{L^2}{d^2}$.

Thus, to go from $n_z = 1$ to $n_z = 2$, it is necessary that $n_x^2 + n_y^2 \geq \frac{3}{4}\frac{L^2}{d^2}$.

5)

a) Taking the potential energy origin to be the vacuum level, the energies in the direction *Tz* are spread as indicated in Figure 4.12.

In direction *Oz*, the energy levels are quantified and the discrete levels correspond to those of a potential well, as shown in the figure above for levels E_{z1}, E_{z2}, and E_{z3}. Due to the energy difference between conduction bands, the electrons in the conduction band of SC_2 accumulate in the potential wells formed in the conduction band of SC_1. If SC_2 is doped, this number of electrons is high while SC_1 is highly degenerate without even being doped. In other words, there is a high number of electrons in the conduction band so that the semiconductor degenerates towards a metallic state. The spatial separation of a high number of electrons into SC_1 as diffusion centers (as many as there are doping agents in SC_2) favors a high mobility of electrons in SC_1.

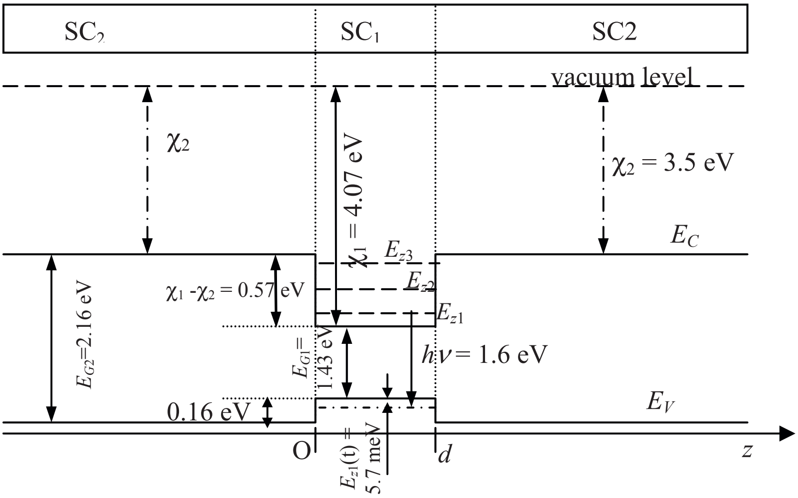


Figure 4.13. *Energy levels in the structure AlAs/Ga As/AlAs*

b) The potential height of the interface between SC_2 and SC_1 is:

$$W_0 = \chi_1(\text{GaAs}) - \chi_2(\text{AlAs}) = 0.57 \text{ eV}.$$

In the potential wells, along *Oz* we can place the E_z levels (relative to the base of the conduction band):

$$E_{z1} = 0.057 \text{ eV},$$
$$E_{z2} = 0.23 \text{ eV},$$
$$E_{z3} = 0.52 \text{ eV}.$$

Finally, the approximation of an infinite height for the potential wells previously used to determine wave functions and energies is acceptable when electrons are situated on the level denoted E_{z1}; however, it is no longer acceptable for level E_{z3}.

For holes, the barrier is given by:

$$(\chi_2 + E_{G2}) - (\chi_1 + E_{G1}) = 5.66 - 5.50 = 0.16 \text{ eV}.$$

c) With respect to the summit of the GaAs valence band (with $E_V = 0$), the position of the hole accepting layer is given by:

$$E_{z1}(t) = \left| E_V - \frac{\hbar^2}{2m_t^*} k_z^2 \right| = \left| E_V - \frac{\hbar^2}{2m_t^*} \frac{\pi^2}{d^2} n_z^2 \right| \overset{\substack{E_V = 0 \\ n_z = 1}}{=} \frac{\hbar^2}{2\left|m_t^*\right|} \frac{\pi^2}{d^2} \simeq 5.7 \text{ meV}.$$

d) The transition has a frequency ν such that:

$$h\nu = E_{z1} + E_{G1} + E_{z1}(t) \simeq 1.6 \text{ eV}.$$

We should note that E_{z1}, like $E_{z1}(t)$, can be adjusted by a change in the value of d (width of the GaAs slice). This makes it possible to change the frequency (and thus the wavelength) of the emission.

We can also note that the frequency that is emitted is as high as that emitted by GaAs alone. In addition, as the extrema of the CB and VB bands are localized on the same SC_1 (GaAs), recombinations are direct and very easy, and this it what defines a so-called type I heterostructure. They find applications in devices such as lasers, due to their high efficiency in photon emission. They were first described by L. Esaki, a winner of the Nobel Prize for physics in 1973.

The deposition of such structures in multi-layer structures, with a periodic repetition along *Tz*, introduces a supplementary periodicity that spreads the E_z levels into bands of electrons or holes.

4.6. Problems

4.6.1. *Problem 1: horizontal tangent at the zone limit (k ≈ π/a) of the dispersion curve*

Using the relation $E=f(k)$ obtained in Chapter 3 for $k \approx \frac{\pi}{a}$ (equation [3.16]), show by calculation how the energy curves $E=f(k)$ give a horizontal tangent at the points where the energy becomes discontinuous.

Answers

The theory for semi-free electrons in a 1D medium (section 3.2.4) makes it possible to state that for values of k close to $\frac{\pi}{a}$ (coefficient A_1 being large) there is a double condition on the value of the energy that equation [3.16] translates as:

$$\left(\frac{2\mu E}{\hbar^2} - k^2\right)\left(\frac{2\mu E}{\hbar^2} - [k - \omega]^2\right) = \frac{\mu^2 w_0^2}{\hbar^4}$$

where $\omega = \frac{2\pi}{a}$.

The left-hand side member is a function of E and k, while $g(E, k)$ is such that $g(E,k) = \frac{\mu^2 w_0^2}{\hbar^4}$. If we differentiate this equation, we obtain $g'_E dE + g'_k dk = 0$, or rather $\frac{dE}{dk} = -\frac{g'_k}{g'_E}$. As we wish to show that $\left[\frac{dE}{dk}\right]_{k=\frac{\pi}{a}} = 0$, it suffices from the preceding equation to show that $[g'_k]_{k=\frac{\pi}{a}} = 0$. So, calculating g'_k gives us:

$$g'_k = \frac{\partial g}{\partial k} = -2k\left[\frac{2\mu E}{\hbar^2} - \left(k - \frac{2\pi}{a}\right)^2\right] + \left[\frac{2\mu E}{\hbar^2} - k^2\right]\left[-2\left(k - \frac{2\pi}{a}\right)\right],$$

from which

$$\left[g'_k\right]_{k=\frac{\pi}{a}} = -\frac{2\pi}{a}\left[\frac{2\mu E}{\hbar^2} - \frac{\pi^2}{a^2}\right] + \left[\frac{2\mu E}{\hbar^2} - \frac{\pi^2}{a^2}\right]\left[-2\left(-\frac{\pi}{a}\right)\right] = 0,$$

which is the answer required.

4.6.2. *Problem 2: scale of m* in the neighborhood of energy discontinuities*

1) The preceding sections detailed two conditions on energy for semi-free electrons placed in a potential defined by $V(x) = w_0 \cos\omega x$. Now show that:

$$E = \frac{1}{2}\left[E_k^0 + E_{k-\omega}^0 \pm \sqrt{\left(E_k^0 - E_{k-\omega}^0\right)^2 + w_0^2}\right].$$

2) In the neighborhood of a discontinuity, we can state that $k = \frac{\pi}{a} - \Delta k$ where $\Delta k << \frac{\pi}{a}$ and $\Delta k > 0$ or $\Delta k < 0$. By making $E_B^0 = \frac{\hbar^2}{2m}\left(\frac{\pi}{a}\right)^2$, show that:

$$E - E_B^0 \simeq \frac{\hbar^2}{2m}\Delta k^2\left[1 \pm \frac{4E_B^0}{w_0}\right] \pm \frac{w_0}{2}.$$

3) Give the limiting values for E for a discontinuity obtained for $\mathrm{k} = \frac{\pi}{\mathrm{a}}$. From this deduce the size of the forbidden band.

4) In this problem, $|w_0| = 2$ eV and $a = 0.3$ nm. Calculate the ratio $\frac{m^*}{m}$ for the bottom of the second band and for the top of the first band.

Answers

1) The theory for semi-free electrons in a 1D medium makes it possible to state that for values of k close to $\frac{\pi}{a}$ there is a double condition on the energy value. The compatibility of these two conditions brings us to equation [3.16]:

$$\left(E - \frac{\hbar^2 k^2}{2\mu}\right)\left(E - \frac{\hbar^2[k - \omega]^2}{2\mu}\right) = \frac{w_0^2}{4}.$$

By making $E_k^0 = \frac{\hbar^2 k^2}{2\mu}$ and $E_{k-\omega}^0 = \frac{\hbar^2 (k-\omega)^2}{2\mu}$ (where $\omega = \frac{2\pi}{a}$), this equation can also be written as $(E - E_k^0)(E - E_{k-\omega}^0) = \frac{w_0^2}{4}$, which is a second degree equation:

$$E^2 - E(E_k^0 + E_{k-\omega}^0) + \left(E_k^0 E_{k-\omega}^0 - \frac{w_0^2}{4} \right) = 0.$$

Its solution is written as:

$$E = \frac{1}{2}\left[E_k^0 + E_{k-\omega}^0 \pm \sqrt{\left(E_k^0 + E_{k-\omega}^0\right)^2 - 4\left(E_k^0 E_{k-\omega}^0 - \frac{w_0^2}{4} \right)} \right],$$

which is also written as:

$$E = \frac{1}{2}\left[E_k^0 + E_{k-\omega}^0 \pm \sqrt{\left(E_k^0 - E_{k-\omega}^0\right)^2 + w_0^2} \right].$$

2) The preceding equation is for energy in the neighborhood of $k \simeq \frac{\pi}{a} = \frac{\omega}{2}$.

Making $k = \frac{\pi}{a} - \Delta k = \frac{\omega}{2} - \Delta k$, where $\Delta k > 0$ at the top of the band, and $\Delta k < 0$ at the bottom of the band, and in addition assuming that $\Delta k << \frac{\pi}{a}$, we have for values of $k = \frac{\omega}{2} - \Delta k$ the equations $E_k^0 = \frac{\hbar^2}{2\mu}\left(\frac{\omega^2}{4} - \omega\Delta k + \Delta k^2 \right)$ and $E_{k-\omega}^0 = \frac{\hbar^2}{2\mu}\left(\frac{\omega^2}{4} + \omega\Delta k + \Delta k^2 \right)$.

Under these conditions the preceding equation for E can be written as:

$$E = \frac{1}{2}\left\{ \frac{\hbar^2}{2m}\left[\frac{\omega^2}{2} + 2\Delta k^2 \right] \pm w_0 \left[1 + \left(\frac{\hbar^2}{2m} \right)^2 \frac{4\omega^2 \Delta k^2}{w_0^2} \right]^{1/2} \right\}.$$

By making

$$E_B^0 = \frac{\hbar^2}{2m}\left(\frac{\omega}{2}\right)^2 \overset{\omega=\frac{2\pi}{a}}{=} \frac{\hbar^2}{2m}\left(\frac{\pi}{a}\right)^2 = E^0_{k=\pi/a},$$

the two solutions for energy are such that:

– first solution:

$$\begin{aligned} E_{(1)} - E_B^0 &= \frac{\hbar^2}{2m}\Delta k^2 + \frac{w_0}{2}\left[1 + \frac{\hbar^2}{2m}\left(\frac{\hbar^2}{2m}\frac{\omega^2}{4}\frac{4.4\Delta k^2}{w_0^2}\right)\right]^{\frac{1}{2}} \\ &= \frac{\hbar^2}{2m}\Delta k^2 + \frac{w_0}{2}\left[1 + \frac{\hbar^2}{2m}\left(E_B^0\frac{4.4\Delta k^2}{w_0^2}\right)\right]^{\frac{1}{2}} \\ &\simeq \frac{\hbar^2}{2m}\Delta k^2\left[1 + \frac{4E_B^0}{w_0}\right] + \frac{w_0}{2} \end{aligned}$$

– second solution:

$$E_{(2)} - E_B^0 \simeq \frac{\hbar^2}{2m}\Delta k^2\left[1 - \frac{4E_B^0}{w_0}\right] - \frac{w_0}{2}.$$

Both solutions can be brought together in the equation written:

$$E - E_B^0 \simeq \frac{\hbar^2}{2m}\Delta k^2\left[1 \pm \frac{4E_B^0}{w_0}\right] \pm \frac{w_0}{2}.$$

3) At the limit, as $\Delta k \to 0$:

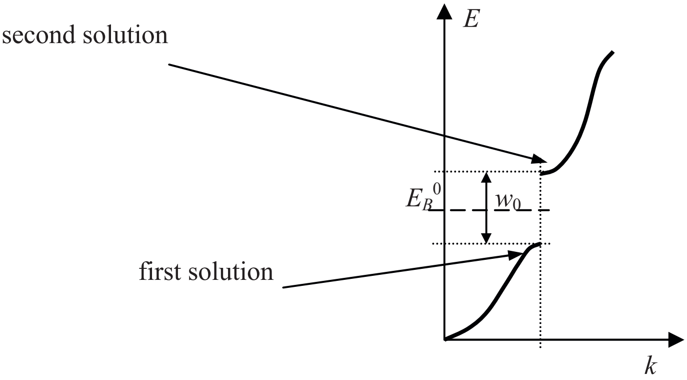
– the second solution goes towards

$$E_{(2)} = E_B^0 - \frac{w_0}{2}.$$

With $w_0 < 0$, this solution is for the bottom of the band where the value of Δk introduced into the initial expression, $k = \frac{\pi}{a} - \Delta k$, must be negative.

– the first solution tends towards $E_{(1)} = E_B^0 + \frac{w_0}{2}$.

With $w_0 < 0$, this solution is for the top of the band where the value of Δk must be positive.



The gap between the two solutions for the energy is equal to $|w_0| = E_G$.

4) To calculate the effective mass at the top of the band (where it must be negative), we will use the expression for E corresponding to the first answer to problem 2, that is:

$$E_{(1)} \simeq E_B^0 + \frac{\hbar^2}{2m}\Delta k^2\left[1 + \frac{4E_B^0}{w_0}\right] + \frac{w_0}{2}.$$

This gives the effective mass as $\frac{1}{m_{(1)}^*} = \frac{1}{\hbar^2}\frac{\partial^2 E_{(1)}}{\partial k^2}$. Because $k = \frac{\omega}{2} - \Delta k$, we can note that

$$\frac{\partial E_{(1)}}{\partial k} = \frac{\partial E_{(1)}}{\partial \Delta k}\frac{\partial \Delta k}{\partial k} = -\frac{\partial E_{(1)}}{\partial \Delta k} \Rightarrow \frac{\partial^2 E_{(1)}}{\partial k^2} = \frac{\partial^2 E_{(1)}}{\partial \Delta k^2}.$$

With

$$\frac{\partial E_{(1)}}{\partial \Delta k} = \frac{\hbar^2}{m}\left(1 + \frac{4E_B^0}{w_0}\right)\Delta k\,,$$

from which

$$\frac{\partial^2 E_{(1)}}{\partial \Delta k^2} = \frac{\hbar^2}{m}\left(1 + \frac{4E_B^0}{w_0}\right),$$

we deduce that:

$$\frac{1}{m^*_{(1)}} = \frac{1}{\hbar^2}\frac{\partial^2 E_{(1)}}{\partial \Delta k^2} = \frac{1}{m}\left(1 + \frac{4E_B^0}{w_0}\right).$$

Numerically, with $w_0 = -2$ eV and $a = 0.3$ nm, we obtain $E_B^0 \simeq 4.2$ eV, from which:

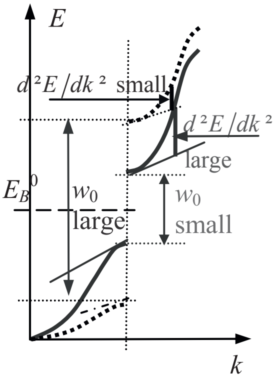
$$\frac{m^*_{(1)}}{m} \simeq -0.135.$$

To calculate the effective mass at the bottom of the second band (where it must be positive) we use the second expression for E found in the answer to problem 2, i.e.:

$$E_{(2)} \simeq E_B^0 + \frac{\hbar^2}{2m}\Delta k^2\left[1 - \frac{4E_B^0}{w_0}\right] - \frac{w_0}{2}.$$

This gives the effective mass as:

$$\frac{1}{m^*_{(2)}} = \frac{1}{\hbar^2}\frac{\partial^2 E_{(2)}}{\partial k^2} = \frac{1}{\hbar^2}\frac{\partial^2 E_{(2)}}{\partial \Delta k^2} = \frac{1}{m}\left(1 - \frac{4E_B^0}{w_0}\right), \text{ so that } m^*_{(2)}\big/m \simeq +0.106.$$



In the figure above, we can note that if w_0 is large (strong interaction with the lattice) the permitted bands are relatively narrow (flat) and d^2E/dk^2 is small. Close to the discontinuity, $m^* \propto \left(d^2E/dk^2\right)^{-1}$ is thus just as large (or small) as the interaction is strong (or weak). When k increases (on the right of the discontinuity) d^2E/dk^2 decreases and m^* increase up to *a* infinite value in the middle of the permitted band (m^* is at a minimum at the base of the band).

4.6.3. *Problem 3: study of $E_F(T)$*

We showed in Chapter 2, problem 4 that Fermi level of a 3D metal was practically independent of temperature, assuming the conduction electrons are free electrons. Is this property still valid if the electrons of the metal are assumed to be semi-free?

Answers

For a 3D metal that has a metallic character originating from electrons placed in an incompletely full conduction band (typically half-full as indicated in Chapter 6), the state density function [$Z(E)$] takes on the same form for semi-free electrons as for free electrons. Only the $\sqrt{|E - E_n|}$ term must be replaced by the term $\sqrt{E}$, where $E_n = E_C$ is the potential energy of the bottom of the conduction band. The other functions (Fermi-Dirac notably) remain invariable. Simplifying, so as to not have to go through the calculations again, we can suppose that a judicious choice of origins at the bottom of the conduction band will suffice (so that $E_n = E_C = 0$), as this choice is the same as fixing the value of $\sqrt{|E - E_n|}$ to that of $\sqrt{E}$. $Z(E)$ thus takes on exactly the same form as for the free electrons as long as the mass of the electron is replaced by the effective mass of the semi-free electron. The property of *quasi*-invariability of the Fermi level with temperature is thus retained when going from using free to semi-free electrons to represent metal electrons.

Chapter 5

Crystalline Structure, Reciprocal Lattices and Brillouin Zones

5.1. Periodic lattices

5.1.1. *Definitions: direct lattice*

A lattice is a periodic (hence regular) arrangement of points called nodes. At each node the base of an atom is attached, the nature of which depends on the solid.

In 3D, the lattice is defined by three fundamental vectors denoted $\vec{a}, \vec{b}, \vec{c}$ that are such that the atomic arrangement is identical around a point *P* defined with respect to the origin *O* by $\vec{r} = \overrightarrow{OP}$, and around a point *P*' defined by $\overrightarrow{r'} = \overrightarrow{OP'}$. The vector $\overrightarrow{r'}$ must be such that $\overrightarrow{r'} = \vec{r} + \vec{T}$ where $\vec{T}$ is a translational vector defined by $\vec{T} = n_1\vec{a} + n_2\vec{b} + n_3\vec{c}$, and where n_1, n_2 and n_3 are integers.

Furthermore, we can see that the lattice, which is characteristic of crystalline structures, is created by the addition (superposition) of the base (ellipsoid) to each node (point) of the crystal.

Thus, any two points of a lattice are always linked to one another by an appropriate translational vector $\vec{T}$.

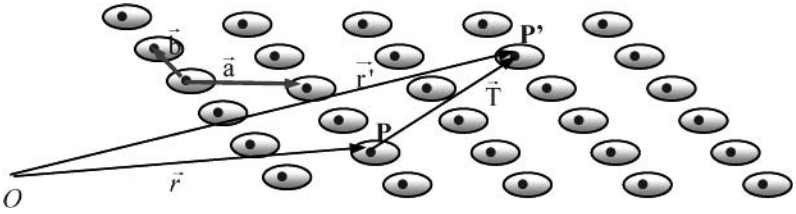


Figure 5.1. *Imaginary crystal element where nodes are represented by points encircled by an ellipsoid, representing the base of an atom*

In the 2D figure above we can see that the atomic environment is identical around points P and P' with each being linked to one another by the vector $\vec{T} = n_1\vec{a} + n_2\vec{b}$ which is such that $n_1 = 2$ and $n_2 = 3$. The vectors $\vec{a}$ and $\vec{b}$ shown in the figure are fundamental vectors.

The parallelepiped defined by the fundamental vectors $\vec{a}, \vec{b}, \vec{c}$ is called the primitive unit cell, and is the smallest unit possible. The volume of the primitive unit cell is equal to $V_c = |\vec{a} \cdot (\vec{b} \times \vec{c})|$. On average, each parallelepiped unit cell contains one node, and each node is shared with four cells below and four cells above the base plane of the node itself. In effect, each node is shared between eight cells and as there are eight nodes per cell (at the eight tops of the parallelepiped) with each being shared with eight other cells, we have an average of one node per cell.

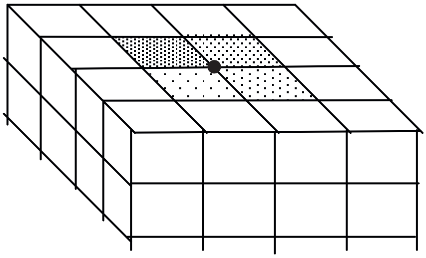


Figure 5.2. *Parallepipedic structure with an average one node per cell*

The various crystalline structures belong to different types of lattice. In all there are fourteen Bravais lattices (or "space lattices"), with the simplest structures being simple cubic (sc), body-centered cubic (cc), face-centered cubic (fcc), tetragonal, orthorhombic, hexagonal, triclinic, etc. (see any good book on crystalline structures for futher details).

5.1.2. *Wigner-Seitz cell*

There is another type of primary cell with a volume equal to $V_c = |\vec{a} \cdot (\vec{b} \times \vec{c})|$. It is the Wigner-Seitz cell that is found in the following manner. As shown in Figure 5.3:

– the lines between a given node and its neighbors are traced;

– another line (or plane in the case of 2D lattices) is drawn normal to the first lines at their mid-points; and

– the volume (or the smallest surface) enclosed by these lines (or plane) is the Wigner-Seitz primary cell. This cell can cover or fill the whole space of the lattice.

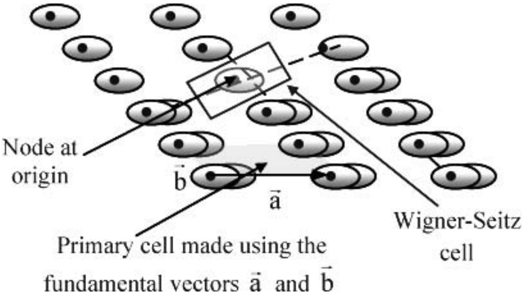


Figure 5.3. *Wigner-Seitz cell*

5.2. Locating reciprocal planes

5.2.1. *Reciprocal planes: definitions and properties*

All nodes of a lattice can be grouped into classes of parallel and equidistant planes called reticular planes (there are in reality an infinite number of reticular planes). The position and orientation of a reticular plane is determined by any three points (nodes) in this plane. If the base contains only one atom, then the plane contains three unaligned atoms. These nodes can, in certain cases, determine the so-called cleavage plane.

5.2.2. *Reciprocal planes: location using Miller indices*

5.2.2.1. *Definition of indices*

If each of the three nodes that define the reciprocal plane are on an axis that carries the fundamental vectors $\vec{a}, \vec{b}, \vec{c}$, then the reciprocal plane can be determined using the abscissae of the nodes with respect to the axis. So if, for example, the three nodes defining the reciprocal plane have the coordinates (with respect to the origin of the fundamental vector) (m, 0, 0), (0, n, 0) and (0, 0, p), then the plane can be found using the respective numbers indicated by m, n, and p. An example is shown in Figure 5.5a using the nodes M, N and P.

If a reciprocal plane is parallel to one of the fundamental vectors, then the preceding numbers are infinite (see Figures 5.5b and 5.5d). To remove this slight inconvenience, we can use Haüy indices, which are the inverse of the preceding numbers. The result is given in simple fractions (that Haüy termed rational indices) that are rather difficult to manipulate. Because of this, Miller proposed the use of whole indices obtained from the multiplication of simple fractions $\frac{1}{m},\frac{1}{n},\frac{1}{p}$ by the same number K. K should be as small as possible so that the result is a succession of integers (hkl). These numbers are called Miller indices, and are such that:

$$h = K\frac{1}{m},\ k = K\frac{1}{n},\ l = K\frac{1}{p}.$$

For example, the coordinates of the point of intersection between a reciprocal plane and the axis of the fundamental vectors $\vec{a},\vec{b},\vec{c}$ are: (4, 0, 0), (0, 1, 0) and (0, 0, 2). With $K=4$, we have:

$$\frac{1}{4}\times 4 = 1 = h;\ 1\times 4 = 4 = k;\ \frac{1}{2}\times 4 = 2 = l.$$

The resulting Miller indices are written in brackets, i.e. (142).

If a reciprocal plane cuts through the negative part of the axis, then the corresponding index is negative. This is indicated by writing the negative sign above the index concerned, as in $(hk\bar{l})$ for an intersection between the reciprocal plane and the axis carrying $\vec{c}$ (see Figure 5.5c).

The directions in a crystal are also defined by the three integers h, k and l, but this time using square brackets, as in [hkl]. They are also such that they are the smallest possible integers proportional to the directing cosines of the angles formed between the direction considered and the cell axis. The direction given by [hkl] is only perpendicular to the planes defined by Miller indices in very specific cases, such as when involving the cubic structure.

5.2.2.2. *Equation for reciprocal planes located with Miller indices*

Figure 5.4 shows a system of axes and a plane (π). *OH* is the line that passes through the origin *O* and is normal to the plane (π). $\vec{u}$ is the unit vector in the direction *OX* that carries *OH*. The vector has components (cosine directors) denoted α, β, γ.

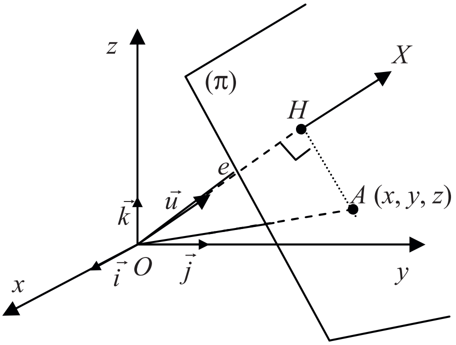


Figure 5.4. *Location of the plane (π)*

For a point (*A*) (with coordinates *x*, *y*, *z*) to belong to the plane (π), it suffices that its projection on the axis *OH* is at *H*. Thus, $\vec{u} \bullet \overrightarrow{OA} = \overline{OH}$, and by making $\overline{OH} = e$, we obtain:

$$\alpha x + \beta y + \gamma z = e,$$

which can be rewritten as:

$$\frac{\alpha}{\mathrm{e}} x + \frac{\beta}{e} y + \frac{\gamma}{e} = 1.$$

If *A* → *M* (*m*, 0, 0), where *M* is the point of intersection between (π) and *Ox*, the preceding equation for the plane where $y = z = 0$ gives $\frac{\alpha}{e} m = 1$, so that $\frac{\alpha}{e} = \frac{1}{m}$. Similarly, by making *A* tend towards *N* (0, *n*, 0) and *P* (0, 0, *p*) we can successively deduce that $\frac{\beta}{e} = \frac{1}{n}$ and $\frac{\gamma}{e} = \frac{1}{p}$. On substituting these values into the equation for the plane (π), we now obtain:

$$\frac{x}{m} + \frac{y}{n} + \frac{z}{p} = 1.$$

The Miller indices of the plane (π) are: $h = K\frac{1}{m}, k = K\frac{1}{n}, l = K\frac{1}{p}$, and the equation for the plane (*hkl*) is thus given by: $hx + ky + lz = K$.

5.2.2.3. *Examples using the simple cubic system*

Figure 5.4 shows four examples of a simple cubic structure with its side length equal to *a* and with different reciprocal planes located by Miller indices.

5.3. Conditions for maximum diffusion by a crystal (Laue conditions)

5.3.1. *Problem parameters*

A crystal can be seen as made up of nodes that carry charges (being attached to atoms or ions) capable of re-emitting (more often called diffusing) an incident wave in any direction. Here we are concerned with discovering the directions in which diffusion is at a maximum.

If (ω, $\vec{k}$) and (ω', $\vec{k}'$) are the angular frequency and wave vector of an incident wave and a diffused wave, respectively, then for a crystal that is assumed to be linear we can state that $\omega = \omega'$ so that $|\vec{k}| = |\vec{k}'|$. Here we have an elastic diffusion for which the incident wave's quantum energy conservation, meaning that $\hbar\omega = \hbar\omega'$ so that $\omega = \omega'$ and $k = k'$. We will thus look at the form of the diffused wave which is located with the vector $\vec{\rho}_{m,n,p} = m \cdot \vec{a} + n \cdot \vec{b} + p \cdot \vec{c}$. We will then deduce the conditions required for maximum diffusion.

NOTE.– Diffusion and diffraction

When a crystal is irradiated (with X-rays in crystallography) each charge diffuses the rays as if it were the source diffusing in all directions. All the diffused rays interfere with one another, resulting in a cancelling out of some rays in certain directions (destructive interference) and reinforcement in other directions. The latter gives rise to diffracted rays. In other words, diffraction corresponds to the maximum resultant diffusion.

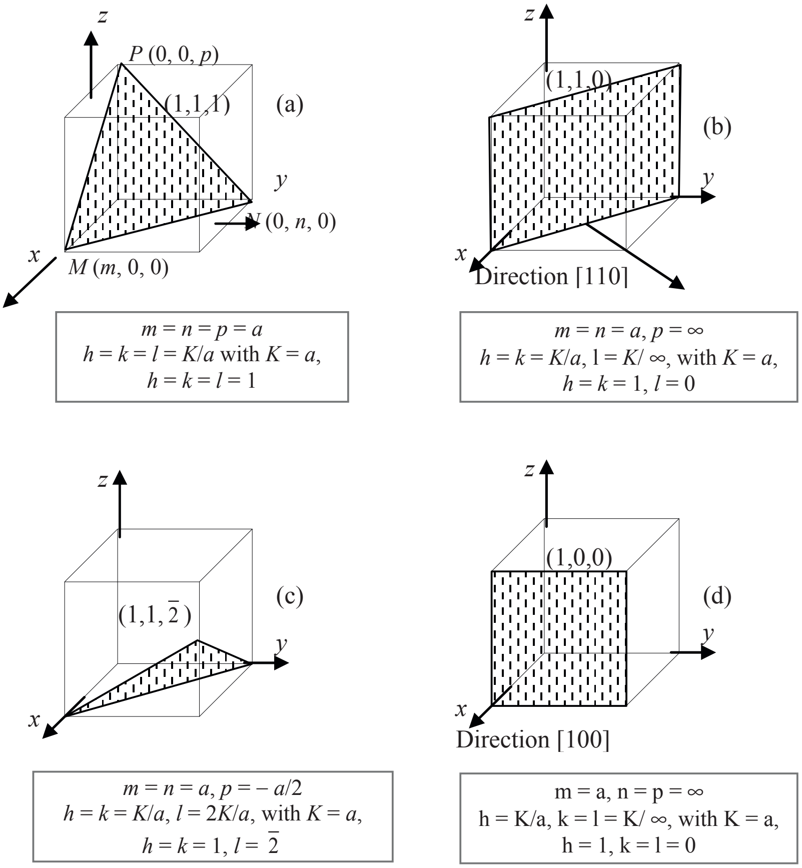


Figure 5.5. *Examples of Miller indices determination in a cubic structure (cube of side a)*

5.3.2. *Wave diffused by a node located by* $\vec{\rho}_{m,n,p} = m \cdot \vec{a} + n \cdot \vec{b} + p \cdot \vec{c}$

5.3.2.1. *Ray diffused by charged particles: a reminder*

Here we consider a simple model for the elastic diffusion of a monochromatic electromagnetic wave that has the form $\vec{E} = \vec{E}_0 \exp i\,(\vec{k} \cdot \vec{r} - \omega t) = \underline{\vec{E}}_0 \exp(-j\,\omega t)$ by a charge that is initially in a relaxed state and is denoted in q. Neglecting frictional and steric effects, q is accelerated by the incident wave as described by the equation $\vec{F} = m^*\vec{\gamma} = q\vec{E}$. At a point *M*, placed at a great distance *r* from *q*, the incident wave diffused by *q* gives rise to a single contribution E_θ along $\vec{e}_\theta$, as in $E_\theta = \frac{q\underline{\gamma}}{4\pi\varepsilon_0 c^2}\frac{\sin\theta}{r}e^{jkr}e^{-j\omega t}$. This can be rewritten as $E_\theta = \underline{E}_{0\theta}e^{-j\omega t}$ where $\underline{E}_{0\theta} = \frac{q\underline{\gamma}}{4\pi\varepsilon_0 c^2}\frac{\sin\theta}{r}e^{jkr}$. If we substitute $\underline{\gamma} = \frac{q}{m^*}\underline{E}_0$ into this equation, then we end up with:

$$\underline{E}_{0\theta} = \frac{1}{4\pi\varepsilon_0 c^2}\frac{q^2\sin\theta}{m^*}\frac{\underline{E}_0}{r}e^{jkr} = C\frac{\underline{E}_0}{r}e^{jkr}. \qquad [5.1]$$

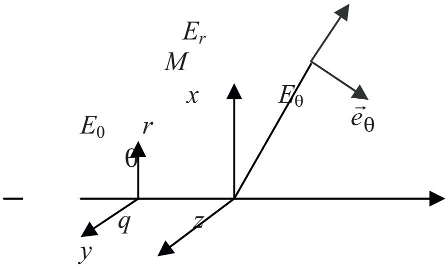


Figure 5.6. *Incident wave diffusion by the charge denoted q*

5.3.2.2. *Determination of the diffraction conditions for a crystal*

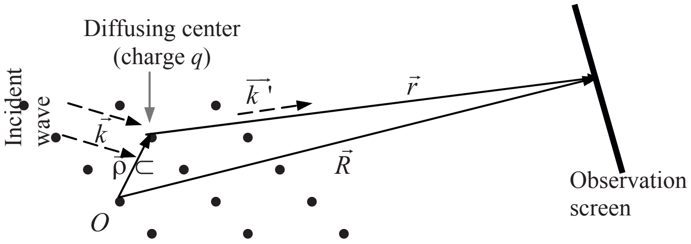


Figure 5.7. *Electromagnetic wave diffusion by a crystal*

Here, the equation for the plane incident wave (with wave vector $\vec{k}$) at a point determined with respect to *O* by the vector $\vec{\rho}$ is given as $\vec{E}_\rho = \underline{\vec{E}}_0\exp(-j\omega t) = \vec{E}_0\exp j(\vec{k}\vec{\rho} - \omega t)$. In accordance with theory, the wave is diffused by charges in the lattice and propagates in the direction given by the wave vector $\overrightarrow{k}'$, where $\left|\vec{k}\right| = \left|\vec{k}'\right|$. From a given observation point, defined by vector $\vec{r}$ with respect to the diffusion point, the complex amplitude of the diffused wave is, from equation [5.1], given by:

$$\underline{E}_{\text{diff}} = C\frac{\underline{E}_0}{r}\exp j\left(\overrightarrow{k'}\cdot\vec{r}\right) = C\frac{\underline{E}_0}{r}\exp j\left(\vec{k}\cdot\vec{\rho}\right)\exp\left(jkr\right),$$

and

$$E_{\text{diff}} = \underline{E}_{\text{diff}}e^{-j\omega t} = CE_0\exp j\left(\vec{k}\cdot\rho\right)\frac{\exp j\left(kr-\omega t\right)}{r}. \qquad [5.2]$$

where $\vec{r} \parallel \vec{k}'$ and $\overrightarrow{k'}\vec{r} = k'r = kr$.

In addition, as we assume that the observation is made at a great distance from crystals, with respect to the size of the crystal, we can state that: $r \simeq R - \rho\cos(\vec{\rho},\vec{R})$. In effect, as the angle given by $(\vec{r},\vec{R})$ is small, we have:

$$\operatorname{proj}_R r = r\cos\left(\vec{r},\vec{R}\right) \simeq r \text{ and } \cos\left(\vec{\rho},\vec{R}\right) = \frac{R-\operatorname{proj}_R r}{\rho} \simeq \frac{R-r}{\rho}$$

where $\operatorname{proj}_R r$ stands for the projection of r on R.

Equation [5.2] for the diffused wave can also be written as $E_{\text{diff}} = CE_0\frac{\exp j(\varphi)}{r}\exp(-j\omega t)$, is thus such that $\exp(j\varphi) = \exp j\left(\vec{k}\cdot\vec{\rho}+kr\right) = \exp\left(jkR\right)\exp j(\vec{k}\cdot\vec{\rho}-k\rho\cos[\vec{\rho},\vec{R}])$.

With $k\rho\cos\left(\vec{\rho},\vec{R}\right) \simeq k'\rho\cos\left(\vec{\rho},\overrightarrow{k'}\right) = \overrightarrow{k'}\cdot\vec{\rho}$, we can now write that:

$$\exp\left(j\varphi\right) = \exp\left(jkR\right)\exp j\left(\vec{k}\cdot\vec{\rho}-\overrightarrow{k'}\cdot\vec{\rho}\right) = \exp(jkR)\exp\left(-j\vec{\rho}\cdot\overrightarrow{\Delta k}\right)$$

where $\overrightarrow{\Delta k} = \overrightarrow{k'} - \vec{k}$.

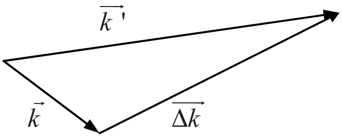


Figure 5.8. *Brillouin zone limits*

Finally, with $r \simeq R$, we can rewrite the wave diffused by a center that is located in a lattice by a positional vector $\vec{\rho}_{m,n,p} = m\vec{a} + n\vec{b} + p\vec{c}$, in the following way:

$$E_{\text{diff}} = \left(\frac{C \cdot E_0 e^{jkR} e^{-j\omega t}}{R} \right) \exp\left(-j \vec{\rho}_{m,n,p} \cdot \overrightarrow{\Delta k} \right).$$

Here the term $\exp\left(-\mathrm{j}\vec{\rho}_{\mathrm{m,n,p}} \cdot \overrightarrow{\Delta \mathrm{k}}\right)$ is the form factor.

To conclude, the total wave diffused in a direction determined by the lattice is the sum over all diffused waves. This sum extends to all lattice nodes denoted by vectors $\vec{\rho}_{m,n,p} = m\vec{a} + n\vec{b} + p\vec{c}$. The generalized form factor denoted A is defined by considering all lattice nodes, as in $A = \sum_{m,n,p} \exp(-j\vec{\rho}_{m,n,p} \cdot \overrightarrow{\Delta k})$ which is the form factor characterising the lattice. The amplitude of the diffusion thus goes through a maximum when each term of the sum is equal to one, i.e.:

$$\vec{\rho}_{m,n,p} \cdot \overrightarrow{\Delta k} = \left(m\vec{a} + n\vec{b} + p\vec{c} \right) \cdot \overrightarrow{\Delta k} = 2\pi\,(n)$$

where n is an integer.

As $\exp(-j\ \vec{\rho}_{m,n,p} \cdot \overrightarrow{\Delta k}) = \exp(-j\ m\vec{a} \cdot \overrightarrow{\Delta k})\exp(-j\ n\vec{b} \cdot \overrightarrow{\Delta k})\exp(-j\ p\vec{c} \cdot \overrightarrow{\Delta k})$, the condition for maximum diffusion is given when q, r, and s are whole numbers and the three following equations are simultaneously satisfied.

$$\vec{a} \cdot \overrightarrow{\Delta k} = 2\pi q \quad \vec{b} \cdot \overrightarrow{\Delta k} = 2\pi r \quad \vec{c} \cdot \overrightarrow{\Delta k} = 2\pi s \tag{5.3}$$

These are called Laue equations and define the point of maximum diffusion. They make it possible to determine the vector $\overrightarrow{\Delta k}$, which combined with $\overrightarrow{k'} = \vec{k} + \overrightarrow{\Delta k}$, determines the directions of $\overrightarrow{k'}$ for maximum diffusion.

5.4. Reciprocal lattice

5.4.1. *Definition and properties of a reciprocal lattice*

We can write the $\overrightarrow{\Delta k}$ vectors that verify equation [5.3] as:

$$\overrightarrow{\Delta k} = h\vec{A} + k\vec{B} + l\vec{C}, \qquad [5.4]$$

where *h*, *k*, and *l* are natural integers and $\vec{A}, \vec{B}, \vec{C}$ are the vectors that will be determined. To ensure $\overrightarrow{\Delta k}$ verifies Laue conditions in equation [5.3], it suffices that the $\vec{A}, \vec{B}, \vec{C}$ vectors are such that:

$$\begin{array}{lll} \vec{A}\cdot\vec{a} = 2\pi & \vec{B}\cdot\vec{a} = 0 & \vec{C}\cdot\vec{a} = 0 \\ \vec{A}\cdot\vec{b} = 0 & \vec{B}\cdot\vec{b} = 2\pi & \vec{C}\cdot\vec{b} = 0 \\ \vec{A}\cdot\vec{c} = 0 & \vec{B}\cdot\vec{c} = 0 & \vec{C}\cdot\vec{c} = 2\pi. \end{array} \qquad [5.5]$$

From the first column we find that $\vec{A} \perp \vec{b}$ and $\vec{A} \perp \vec{c}$ so that $\vec{A}$ is collinear to the vectorial product of $(\vec{b}\times\vec{c})$. From this we can deduce that $\vec{A} = K(\vec{b}\times\vec{c})$, and then by performing a scalar multiplication of both members with $\vec{a}$ the first condition $\vec{A}\cdot\vec{a} = 2\pi$ gives $K = \frac{2\pi}{\vec{a}(\vec{b}\times\vec{c})}$. From this we can deduce the vector $\vec{A}$ and then by circular permutation the vectors $\vec{B}$ and $\vec{C}$, as in:

$$\vec{A} = 2\pi\frac{\left(\vec{b}\times\vec{c}\right)}{\vec{a}\left(\vec{b}\times\vec{c}\right)},\ \vec{B} = 2\pi\frac{\vec{c}\times\vec{a}}{\vec{a}\left(\vec{b}\times\vec{c}\right)} \text{ and } \vec{C} = 2\pi\frac{\left(\vec{a}\times\vec{b}\right)}{\vec{a}\left(\vec{b}\times\vec{c}\right)}. \qquad [5.6]$$

Because the dimensions of these vectors are inverse to length, they are reciprocal vectors. The property that will be defined is such that these vectors will be considered as (base) fundamental to the reciprocal lattice. It is worth remembering that the vectors $\vec{A},\vec{B},\vec{C}$ are only mutually orthogonal when $\vec{a},\vec{b},\vec{c}$ are also mutually orthogonal.

Just as in a direct lattice (DL), the position of the nodes are indicated by the tips of the vectors $\vec{\rho}_{m,n,p} = m\vec{a} + n\vec{b} + p\vec{c}$ (where $[\vec{a},\vec{b},\vec{c}]$ are the base vectors of the DL). Similarly, the vectors $\vec{G}_{h,k,l} = h\vec{A} + k\vec{B} + l\vec{C}$ (where $[\vec{A},\vec{B},\vec{C}]$ should therefore also be the same base vectors of the reciprocal lattice) describe the reciprocal lattice nodes.

The result is that for the reciprocal lattice given by

$$\vec{G}_{h,k,l} = h\vec{A} + k\vec{B} + l\vec{C} \qquad [5.7]$$

the vectors verify the Laue conditions because according to equation [5.4] they are such that:

$$\vec{G}_{h,k,l} = \overrightarrow{\Delta k} \qquad [5.8]$$

where $\overrightarrow{\Delta k}$ follows Laue conditions set out in equation [5.9].

Conversely, and in a practical sense, we can say that if a vector $\overrightarrow{\Delta k}$ is equal to a vector of the reciprocal lattice (given by $\vec{G}_{h,k,l} = h\vec{A} + k\vec{B} + l\vec{C}$) then the vector $\overrightarrow{\Delta k}$ verifies Laue conditions for maximum diffusion (i.e. diffraction).

5.4.2. *Application: Ewald construction of a beam diffracted by a reciprocal lattice*

Here we use the conditions for an elastic diffusion so that there is energy conservation between the incident and diffused waves, i.e. $\hbar\omega = \hbar\omega'$ and $\omega = \omega'$ and $k = k'$. In equation [5.8] we saw that the condition for maximum diffusion for the wave vector is such that $\overrightarrow{\Delta k} = \vec{G}_{h,k,l}$ and in turn $\overrightarrow{\Delta k} = \overrightarrow{k'} - \vec{k}$. Here we thus have:

$$\overrightarrow{k'} = \vec{k} + \vec{G}_{h,k,l}. \qquad [5.9]$$

From the DL we can build the reciprocal lattice $[\vec{A},\vec{B},\vec{C}]$. The vector $\vec{k}$ is then traced parallel to the incident ray and its end point is at a node in the reciprocal lattice. We then draw a sphere (called an Ewald sphere) of radius k and a center of origin $\vec{k}$. The nodes of the reciprocal lattice that belong to this sphere thus make it possible to determine the $\overrightarrow{k'}$ vectors for maximum diffusion. In effect, they are in accordance with both $k = k'$ and $\overrightarrow{k'} = \vec{k} + \vec{G}_{h,k,l}$ as two nodes in the reciprocal lattice belonging to the Ewald sphere are exactly united by the vector $\vec{G}_{h,k,l}$. The diffused ray is thus parallel to $\overrightarrow{k'} = \vec{k} + \vec{G}_{h,k,l}$.

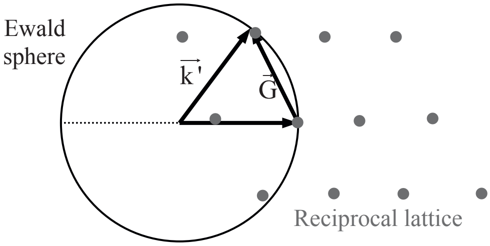


Figure 5.9. *Ewald sphere in a reciprocal lattice*

5.5. Brillouin zones

5.5.1. *Definition*

Briefly, we can define Brillouin zones as being zones delimited by Wigner-Seitz cells in a reciprocal lattice.

5.5.2. *Physical significance of Brillouin zone limits*

Squaring up equation [5.9] for the maximum diffusion of a wave with wave vector $\vec{k}$ gives us:

$$\overrightarrow{k'}^2 = \left(\vec{k} + \vec{G}_{h,k,l}\right)^2 = \vec{k}^2 + \vec{G}_{h,k,l}^2 + 2\vec{k} \cdot \vec{G}_{h,k,l},$$

so that with $k = k'$ (linear crystal):

$$2\vec{k} \cdot \vec{G}_{h,k,l} + \vec{G}_{h,k,l}^2 = 0. \qquad [5.10]$$

This relation can also be written for vector $-\vec{G}_{h,k,l}$ because if $\vec{G}_{h,k,l}$ (as defined by equation [5.7]) is a vector of the reciprocal lattice, then $-\vec{G}_{h,k,l}$ is also a vector of the reciprocal lattice (as it is also in accordance with equation [5.7]).

Equation [5.10] thus gives:

$$2\vec{k}\cdot\vec{G}_{h,k,l}=\vec{G}^2_{h,k,l}, \qquad [5.11]$$

so that:

$$\vec{k}\cdot\left(\frac{1}{2}\vec{G}_{h,k,l}\right)=\left(\frac{1}{2}\vec{G}_{h,k,l}\right)^2 \qquad [5.12]$$

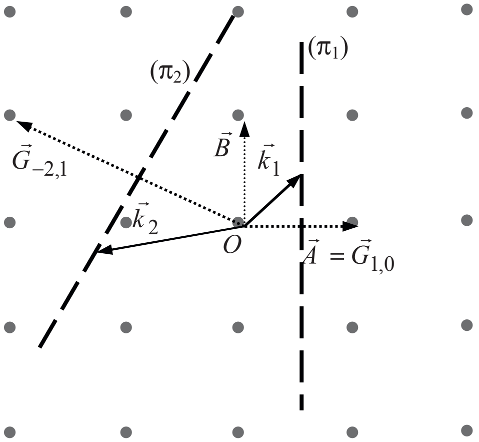


Figure 5.10. *Tracing out Brillouin zone limits*

Figure 5.10 shows a simplified 2D representation (only $\vec{A}$ and $\vec{B}$ vectors are involved) of a reciprocal lattice obtained by taking a node at origin *O* and using the fundamental vectors $\vec{A}$, $\vec{B}$ and $\vec{C}$. The $\vec{G}_{h,k,l}$ vector of the reciprocal lattice joins the origin node with any other of the reciprocal lattice ($\vec{G}_{1,0}=\vec{A}$ and $\vec{G}_{-2,1}=-2\vec{A}+\vec{B}$ in the example given in Figure 5.10) so that the mediating plane (π) is the segment of modulus $\left|\vec{G}_{h,k,l}\right|$. For the system shown in Figure 5.10, π_1 and π_2 are in fact the mediators. As shown in Figure 5.10, $\vec{k}_1$ and $\vec{k}_2$ tie O to π1 and π2, respectively, and indeed all $\vec{k}$ wave vectors going from the origin to this type of mediator plane (π) are in accordance with equation [5.12] which relates $\vec{k}$ to the maximum wave diffusion.

In a regular solid, electrons associated with a wave vector that has an extremity at the plane π (the Brillouin zone limit or the Wigner-Setz cell of the reciprocal lattice) undergoes the maximum diffusion. These electrons are reflected by the DL nodes (ions) and therefore cannot propagate. This is in effect the same property as that found for semi-free electrons of a wave vector that is in accordance with $k = n\frac{\pi}{a}$ at the band limits.

In section 5.6.3 it is thus shown that equation [5.10] from which equation [5.12] is derived is equivalent to the Bragg condition for maximum wave diffusion (diffraction) (presented in section 3.4.2).

5.5.3. *Successive Brillouin zones*

By definition, the first Brillouin zone is the smallest volume in 3D, or the smallest surface in 2D, that is generated by mediating planes created by the segments joining the origin of the reciprocal lattice with its first neighboring nodes. The second zone brings in a similar volume with the exclusion of the first Brillouin zone. This process is repeated *ad infinitum*.

5.6. Particular properties

5.6.1. *Properties of $\vec{G}_{h,k,l}$ and relation to the direct lattice*

Here we show that the reciprocal lattice vector $\vec{G}_{h,k,l} = h\vec{A} + k\vec{B} + l\vec{C}$ is normal to the reticular plane (hkl) of the direct lattice and that $\left|\vec{G}_{h,k,l}\right| = \frac{2\pi}{d_{h,k,l}}$, where $d_{h,k,l}$ is the inter-reticular distance between Miller index planes (hkl).

According to the definition of Miller indices, plane (*hkl*) goes through points *M*, *N* and *P* (Figure 5.9) situated in that order on the three axes *x*, *y* and *z* at distances $\frac{K}{h}, \frac{K}{k}, \frac{K}{l}$ from the origin (*O*) of the DL. *K* is a constant whole number. To show that $\vec{G}_{h,k,l}$ is normal to the plane (*h*, *k*, *l*), we need to show that with two vectors in the plane its scalar product is equal to zero. This can be done by using equations [5.5] shown above. Thus we find:

$$\begin{aligned}\vec{G}_{h,k,l}\cdot\overrightarrow{HK} &= \vec{G}_{h,k,l}\left(\overrightarrow{OK}-\overrightarrow{OH}\right)\\ &= (\mathrm{h}\vec{\mathrm{A}}+\mathrm{k}\vec{\mathrm{B}}+\mathrm{l}\vec{\mathrm{C}})\left(\frac{\mathrm{K}}{\mathrm{k}}\vec{b}-\frac{K}{h}\vec{a}\right)\\ &= 2\pi\mathrm{K} - 2\pi\mathrm{K} = 0.\end{aligned}$$

Similarly, we find that $\vec{G}_{h,k,l}\cdot\overrightarrow{KL}=0$, which indicates the perpendicular nature of $\vec{G}_{h,k,l}$ to the plane (*hkl*).

If we now denote the projection of *O* on the plane *MNP* (plane (*hkl*)) as *H*, i.e. $\overrightarrow{OH}$ goes in the same direction as $\vec{G}_{h,k,l}$ given the aforementioned property, and then by making $\left|\overrightarrow{OH}\right| = d_{hkl}$ which is the inter-reticular distance between planes (*hkl*), we can calculate the scalar product $\vec{G}_{h,k,l}\cdot\overrightarrow{OM}$ by remembering that *OH* is the projection of *OM* on $\vec{G}_{h,k,l}$:

$$\left.\begin{aligned}\vec{G}_{h,k,l}\cdot\overrightarrow{OM} &= \left(h\vec{A}+k\vec{B}+l\vec{C}\right)\cdot\frac{K}{h}\vec{a} = 2\pi(K)\\ &= G_{h,k,l}\cdot OH = G_{h,k,l}\cdot d_{hkl}\end{aligned}\right\}\Rightarrow G_{h,k,l} = \frac{2\pi}{d_{hkl}}\left(K\right). \quad [5.13]$$

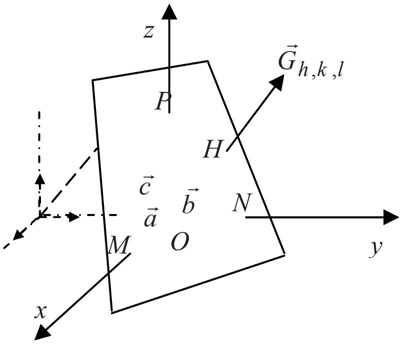


Figure 5.11. *Relation between* $\vec{G}_{h,k,l}$ *and the direct lattice*

5.6.2. *A crystallographic definition of reciprocal lattice*

Crystallographers determining inter-reticular distance generally prefer to use equation [5.14] in place of equation [5.13]. Equation [5.14] relates the reciprocal lattice to the direct lattice more simply. This is because the inter-reticular distance between (*hkl*) planes is simply the inverse of the reciprocal lattice vector modulus (corresponding to $\vec{G}_{h,k,l}$ which is normal to planes (*hkl*)).

$$G_{h,k,l} = \frac{1}{d_{hkl}}(K). \tag{5.14}$$

To obtain this equation, the base vectors of the reciprocal lattice must verify the following equations [5.15] which take the place of equations [5.5]:

$$\begin{array}{lll} \vec{A}\cdot\vec{a} = 1 & \vec{B}\cdot\vec{a} = 0 & \vec{C}\cdot\vec{a} = 0 \\ \vec{A}\cdot\vec{b} = 0 & \vec{B}\cdot\vec{b} = 1 & \vec{C}\cdot\vec{b} = 0 \\ \vec{A}\cdot\vec{c} = 0 & \vec{B}\cdot\vec{c} = 0 & \vec{C}\cdot\vec{c} = 1. \end{array} \tag{5.15}$$

The base vectors defined above by equations [5.6] are now redefined by the following equations [5.16] (which lose their physical significance at the Brillouin zone limits, excepting that the wave vector modulus is given by $k = \frac{1}{\lambda}$ in place of the more normally used $k = \frac{2\pi}{\lambda}$):

$$\vec{A} = \frac{\left(\vec{b}\times\vec{c}\right)}{\vec{a}\left(\vec{b}\times\vec{c}\right)},\ \vec{B} = \frac{\left(\vec{c}\times\vec{a}\right)}{\vec{a}\left(\vec{b}\times\vec{c}\right)} \text{ and } \vec{C} = \frac{\left(\vec{a}\times\vec{b}\right)}{\vec{a}\left(\vec{b}\times\vec{c}\right)}. \tag{5.16}$$

5.6.3. *Equivalence between the condition for maximum diffusion and Bragg's law*

Now let us show that we have an equivalence between the vectorial relation given in equation [5.11], as in $2\vec{k}\cdot\vec{G}_{h,k,l} = \vec{G}^2_{h,k,l}$, and the diffraction condition as given by Bragg (detailed in section 3.4.2). The latter describes the condition for incident rays, reflected by crystal reticular planes, to give constructive interferences.

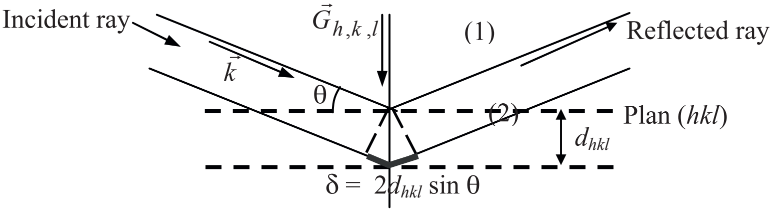


Figure 5.12. *Establishing Bragg's law*

Consider the (*hkl*) planes with an inter-reticular distance denoted by $d_{h,k,l}$ and an incident ray with a wavelength (λ) such that θ is the angle between the wave vector of the incident wave and the (*hkl*) planes. The difference in step between the reflected waves 1 and 2 (shown in Figure 5.12) is such that:

$$\delta = 2d_{hkl}\sin\theta.$$

The constructive interferences (i.e. maximum diffusion points) occur when $\delta = n\lambda$ (*n* being an integer), so that:

$$2d_{hkl}\sin\theta = n\lambda. \qquad [5.17]$$

If we compare equations [5.17] and [5.11] we can see that the latter is of the form:

$$2 \cdot k \cdot G_{h,k,l} \cdot \cos\left(\frac{\pi}{2} - \theta\right) = G^2_{h,k,l}$$

(remember from Figure 5.12 that the vector $\vec{G}_{h,k,l}$ is perpendicular to the (*h*,*k*,*l*) plane so that $\vec{G}_{h,k,l}$ and $\vec{k}$ make an angle equal to $\left(\frac{\pi}{2}-\theta\right)$). Using $k = \frac{2\pi}{\lambda}$, we can rewrite this equation to give:

$$\left.\begin{aligned} G_{h,k,l} &= 2\frac{2\pi}{\lambda}\sin\theta \\ &\overset{(11)}{=} \frac{2\pi}{d_{hkl}}(n), \text{ where } n \text{ is is the integer } K \text{ from equation [5.13]} \end{aligned}\right\} \Rightarrow$$

$$2d_{hkl}\sin\theta = n\lambda,$$

so that in effect we have derived equation [5.17] from equation [5.11]!

NOTE.– In order to be true, equation [5.17] requires that the wavelength λ of the rays should be a similar order of size to that of the distance ($d_{h,k,l}$) between reticular planes. With $d_{h,k,l}$ typically being in the order of tens of nanometers, this means that the incoming rays should be X-rays.

5.7. Example determinations of Brillouin zones and reduced zones

5.7.1. *Example 1: 3D lattice*

To take the simplest example i.e. a simple cubic lattice, we have $|\vec{a}| = |\vec{b}| = |\vec{c}| = a$ where a is the lattice repeat unit. The direct lattice is shown in Figure 5.13.

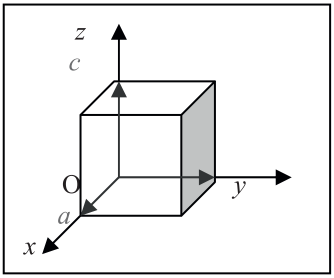


Figure 5.13. *Direct lattice of the simple cubic (sc) structure*

As $\vec{a} \perp \vec{b} \perp \vec{c}$, the defining equation for the base vectors of the reciprocal lattice clearly show that $\vec{A} \perp \vec{B} \perp \vec{C}$.

In addition, with $\vec{a}(\vec{b} \times \vec{c}) = a^3$, i.e. the volume of the primitive unit cell of side a and $\left|\vec{b} \times \vec{c}\right| = a^2$ (vector $\vec{b}$ and $\vec{c}$ being orthogonal and of modulus a), we have:

$$\left|\vec{A}\right| = 2\pi \frac{\left|\vec{b} \times \vec{c}\right|}{\vec{a}(\vec{b} \times \vec{c})} = \frac{2\pi}{a} = \left|\vec{B}\right| = \left|\vec{C}\right|.$$

With the base vectors of the reciprocal lattice being orthogonal and of identical modulus $(\frac{2\pi}{a})$ the reciprocal lattice of the simple cubic lattice also takes on a cubic structure (and with repeat unit $\frac{2\pi}{a}$ in three directions).

The direct application of the Brillouin zones definition shows that the first zone (and those following) is cubic with side $\frac{2\pi}{a}$, and therefore of volume given by:

$$V_{ZB} = \left(\frac{2\pi}{a}\right)^3.$$

The geometric representations for a 2D structure are given in the following section along with a plan figure.

Finally, we can show that for such a structure, we have:

$$d_{hkl} = \frac{a}{\sqrt{h^2 + k^2 + l^2}}. \qquad [5.18]$$

We have $\overrightarrow{G_{h,k,l}}^2 = (h\vec{A} + k\vec{B} + l\vec{C})^2 = h^2A^2 + k^2B^2 + l^2C^2$, as $\vec{a} \perp \vec{b} \perp \vec{c}$ and given the values for $\left|\vec{A}\right|, \left|\vec{B}\right|, \left|\vec{C}\right|$ we obtain $G^2_{h,k,l} = (2\pi)^2\left(\frac{h^2}{a^2} + \frac{k^2}{a^2} + \frac{l^2}{a^2}\right)$.

Using equation [5.13] in the form $d_{h,k,l} = \frac{2\pi}{G_{h,k,l}}$, makes it possible to obtain equation [5.18] as planned.

5.7.2. *Example 2: 2D lattice*

As a simple example of a real plane crystal we can use the square cell of side a (as in $a = b$) shown in Figure 5.14. In the direct lattice, we have for equation [5.15] applied to 2D, as in:

$$d_{hk} = \frac{a}{\sqrt{h^2 + k^2}},$$

so that $d_{11} = \frac{a}{\sqrt{2}}$. In addition, we know that $\vec{G}_{1,1}$ is normal to the reticular plane (see equation [5.13]) with a modulus equal to $2\pi/d_{11,}$ and so that $\left|\overrightarrow{G_{1,1}}\right| = \frac{2\pi}{a}\sqrt{2}$. This can be verified in the Figure shown for the reciprocal lattice (Figure 5.15a).

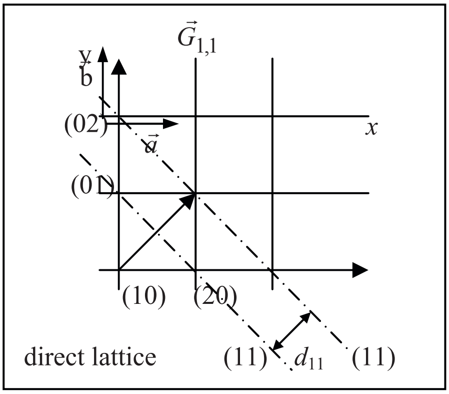
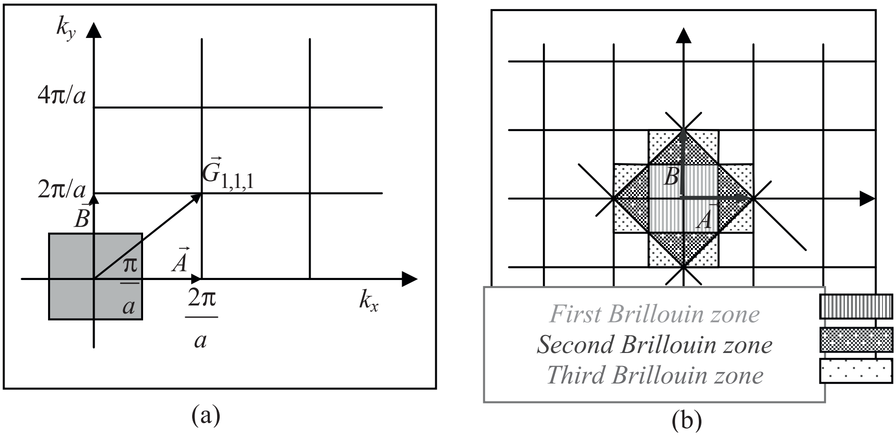


Figure 5.14. *Square direct lattice*

At the level of the reciprocal lattice, the base vector $\vec{A}$ should be such that $\vec{A}\cdot\vec{a} = 2\pi$ and $\vec{A}\cdot\vec{b} = 0$, so that $\vec{A}\perp\vec{b}$ and $\vec{A}//\vec{a}$. Similarly, $\vec{B}\perp\vec{b}$ and in terms of modulus, we have:

$$\left|\vec{A}\right| = \frac{2\pi}{a} \text{ and } \left|\vec{B}\right| = \frac{2\pi}{b} = \frac{2\pi}{a}.$$



(a) (b)

Figure 5.15. *(a) Reciprocal lattice of a square direct lattice; and (b) Brillouin zones*

We can deduce that: $\overrightarrow{G_{11}} = \vec{A} + \vec{B}$, so that $G_{11}{}^{2} = A^{2} + B^{2} = \frac{4\pi^2}{a^2} + \frac{4\pi^2}{a^2}$, and $G_{11} = \frac{2\pi}{a}\sqrt{2} = \frac{2\pi}{d_{11}}$ where $d_{11} = \frac{a}{\sqrt{2}}$. We also have, for example with $d_{2,1} = \frac{a}{\sqrt{5}}$ such that $\left|\overrightarrow{G_{21}}\right| = \left|2\vec{A} + \vec{B}\right| = \frac{2\pi}{d_{21}} = \frac{2\pi\sqrt{5}}{a}$. Given the general definition, the Brillouin zones are represented in terms of the reciprocal lattice, as shown by the first gray zone in Figure 5.15a and the three zones in Figure 5.15b. The first zone is a square with sides equal to $\frac{2\pi}{a}$. The following zones have the same dimensions. We can check this in Figure 5.16 where we have placed the second zone into the first to give a representation of reduced zones as already used for dispersion curves (Figure 4.9).

In order to do this, we simply need to apply a translation of modulus $\left|\vec{A}\right| = \frac{2\pi}{a}$ with a direction and sense appropriate to each of the elements that make up the second zone (2_a, 2_b, 2_c, 2_d) in order to exactly cover the square surface of the first zone (surface of value $\left[\frac{2\pi}{a}\right]^2$). A similar transformation can also be carried out taking the third zone into the first, and so on. It is thus shown that in this case, the Brillouin zones (BZ) are all exactly the same size.

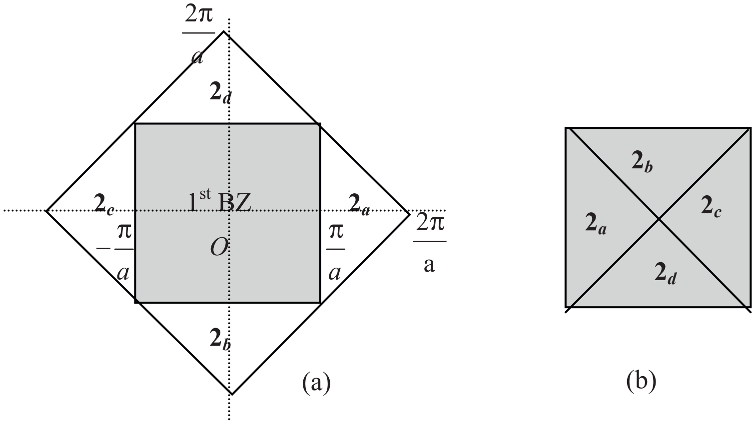


Figure 5.16. *Reduction of the second zone to the first by translation of modulus 2π/a for each element* (2_a, 2_b, 2_c, 2_d)

5.7.3. *Example 3: 1D lattice with lattice repeat unit (a) such that the base vector in the direct lattice is $\vec{a}$*

The direct lattice with the lattice repeat unit denoted by *a* can be simply schematized as below in Figure 5.17.

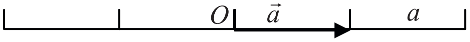


Figure 5.17. *1D periodic lattice*

The fundamental vector $\vec{A}$ of the reciprocal lattice is such that $\vec{A}\cdot\vec{a} = 2\pi$, and its modulus verifies $\left|\vec{A}\right| = \frac{2\pi}{a}$.

The reciprocal lattice along with the successive Brillouin zones are traced in Figure 5.18. We can see without too much difficulty that the various Brillouin zones are of the same size (length 2π/*a*).

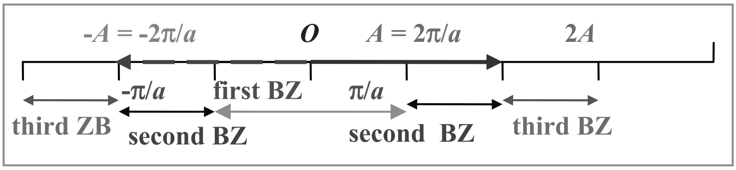


Figure 5.18. *Reciprocal lattice and the Brillouin zone (BZ) of a 1D lattice*

5.8. Importance of the reciprocal lattice and electron filling of Brillouin zones by electrons in insulators, semiconductors and metals

5.8.1. *Benefits of considering electrons in reciprocal lattices*

In Chapter 4 we saw that dispersion curves can take on different forms for the various directions of the $\vec{k}$ vector, and that a plot of $E = f(k_i)$ occurs for each of the i directions that $\vec{k}$ takes on (see in particular Figure 4.4).

Working in the reciprocal lattice obviates the inconvenience of looking for more or less arbitrary directional variations in some parameter (generally electron energy) as we now have a direct expression over all $\vec{k}$ directions, once the reciprocal lattice is plotted. This benefit becomes all the more important given that the Brillouin zones are defined and plotted within the reciprocal lattice for which the zone limits actually have a physical significance. We have seen, significantly, that these limits separate two adjacent zones where electrons can propagate, and where waves (associated with electrons) touch the zone with the extremities of their $\vec{k}$ vectors limit and are stationary, there being two electrons' energy values separated by a gap.

Rather than filling permitted bands with electrons in accordance with the representations given by $E = f(k_i)$ (sufficient for 1D media where there is only one k_x direction to consider), it is interesting to study the electronic filling of successive Brillouin zones so as to obtain precise information on the capacity of each band to accept electrons (capacity being directly comparable over all $\vec{k}$ directions) and on the directions that are permitted for filling.

5.8.2. *Example of electron filling of Brillouin zones in simple structures: determination of behaviors of insulators, semiconductors and metals*

5.8.2.1. *Cubic structure*

5.8.2.1.1. The basics: cells in reciprocal space

Firstly, we can recall that k space can be divided into primitive cells attached to electronic states. For free electrons associated with a progressive wave, we have seen in section 2.6.2.2 that the sides of each primitive cell can be described by $\Delta k_x = \frac{2\pi}{L_1}$, $\Delta k_y = \frac{2\pi}{L_2}$, $\Delta k_z = \frac{2\pi}{L_3}$ so that the volume of each cell is $\Delta k^3 = \Delta k_x \, \Delta k_y \, \Delta k_z = \frac{8\pi^3}{V}$ where $V = L_1 L_2 L_3$ is the volume of a crystal with sides L_1, L_2, L_3 in direct space.

We should also note that the size of the primitive cell is conserved for semi-free electrons. In effect, the wave function is like a Bloch function for which the amplitude changes with *x*, as in: $\psi_{k_x}(x) = e^{ik_x x} u(x)$. The separation of variables means writing the wave function in the form of a product of three functions, each dependent on *x*, *y* and *z*. The condition due to periodic limits (that of Born von Karman) can thus be written with respect to *x* as:

$$\psi_{k_x}(x) = e^{ik_x x} u(x) = \psi_{k_x}(x + L_x) = e^{ik_x (x + L_x)} u(x + L_x).$$

As *u*(*x*) is periodic and because $L_x \simeq N_x a$ if N_x is the number of atoms along *Ox*, then we can deduce that $e^{ik_x L_x} = 1$ and in turn that $k_x = \frac{2\pi}{L_x} n_x$ where n_x is an integer. We therefore find that the partition in *k* space along *Ox* is with cells of dimension $\Delta k_x = \frac{2\pi}{L_x}$, and is identical to that for free electrons (see for example the partition along *k* in Figure 3.2).

5.8.2.1.2. Filling the Brillouin zone of a simple cubic crystal structure

For a cubic crystal at the direct lattice level, and whatever the cell repeat unit in the three directions *x*, *y*, and *z*, the volume of the elementary cell is equal to a^3. The number of primitive cells in the direct lattice of volume denoted *V* is therefore $N = \frac{V}{a^3}$. As each cell contains on average one node, and each node is attached to the base of each atom, the total number of atom bases is also *N*. The reciprocal lattice, as we have seen in section 5.7.1, is also a simple cubic structure, while the first Brillouin zone is a cube of side $\frac{2\pi}{a}$ and therefore of volume $\frac{8\pi^3}{a^3}$. We can therefore place into this Brillouin zone a number of primitive cells equal to:

$$\frac{\frac{8\pi^3}{a^3}}{\Delta k^3} = \frac{8\pi^3}{a^3}\frac{V}{8\pi^3} = \frac{V}{a^3},$$

where $\frac{V}{a^3} = N$ is precisely equal to the number of atom bases in the direct lattice. Given that there is spin, we can place up to two electrons into each primitive cell. This means in turn that we can put two *N* electrons into each Brillouin zone.

So, if in the direct lattice, an atom base liberates two electrons, the first Brillouin zone is just filled by those free electrons leaving the following zone completely empty. The two zones are separated by a "gap" characteristic of the material under consideration. If the gap is in the order of 5 eV or more, then the material is an insulator, but if the value is in the order of 0.5 to 3 eV then it can be thought of as a "moderate" semiconductor.

A bivalent material (based on one atom liberating two electrons) crystallized into a simple cubic structure is therefore an insulator or a semiconductor with a full valence band and an empty conduction band (at absolute zero). The same result would be obtained with a material that had an atomic base containing two monovalent atoms. However, if the material had an atomic base consisting of a single monovalent atom, then the first band would be half-filled so that the electrons could find themselves in free levels and be easily moved under a weak electric field: in other words the material would be a metal.

5.8.2.2. *Filling Brillouin zones in a 2D square structure*

That stated above for the 3D structure can equally be applied to a 2D structure. If S is the surface of a 2D crystal, the number of primitive cells that can be placed in it in a direct lattice is given by: $\frac{S}{a^2} = N_{(2)}$. This number is equal to the number of atom bases in the direct lattice.

The surface of all the Brillouin zones is the same, and is given by $\left(\frac{2\pi}{a}\right)^2$, as we have seen in the scheme of reduced zones (see Figure 5.14). In the Brillouin zones we can place a number of cells equal to:

$$\frac{\frac{4\pi^2}{a^2}}{\Delta k^2} = \frac{4\pi^2}{a^2}\frac{S}{4\pi^2} = \frac{S}{a^2} = N_{(2)}.$$

Once again the number of cells in each Brillouin zone is equal to the number of atom bases. Finally, if each atom base liberates an even number of electrons (N_p) then the first Brillouin zones will be full and the material will be an insulator or a semiconductor. If $N_p = 2$, then only the first Brillouin zone will be full; if $N_p = 4$, then the first two Brillouin zones will be full; if $N_p = 6$ then the first three Brillouin zones will be full; etc.

When N_p is odd, then we have a metal, as in for example $N_p = 5$, and the first two Brillouin zones are full and the third is half full.

5.9. The Fermi surface: construction of surfaces and properties

5.9.1. *Definition*

The Fermi surface is the name given to an equi-energy surface (a constant energy E_F) plotted in wavenumber k space (reciprocal space). Given the Fermi-Dirac function at absolute zero, the Fermi surface separates the occupied electronic states (full orbitals) from the empty states (unfilled orbitals) at this temperature.

5.9.2. *Form of the free electron Fermi surface*

5.9.2.1. *Establishing the form of equi-energy surfaces*

In 3D, free electron energy is given by: $E = \frac{\hbar^2}{2m}(k_x^2 + k_y^2 + k_z^2)$, where for a cubic crystal of length L we have $k_x = k_y = k_z = \frac{2\pi}{L}(n)$. The equi-energy surface $E = E_F$ can be obtained in k space when:

$$k = k_F = \sqrt{(k_x^2 + k_y^2 + k_z^2)_F} = \frac{\sqrt{2mE_F}}{\hbar} \ (= \text{constant}).$$

This surface is thus a sphere with a radius given by $R_{E_F} = k_F$, where k_F is defined by the preceding relation.

For a crystal shaped as a parallelepiped, and with sides given by $L_1 \neq L_2 \neq L_3$ the Fermi surface as given above transforms into an ellipsoid.

5.9.2.2. *The form of Fermi surfaces as found in different representations*

The free electron Fermi surface for a 2D crystal of a square lattice is traced in Figure 5.19a for an arbitrary electron concentration. It can be a hindrance that parts of the Fermi surface belong to the same zone, e.g. the second, and appear separated from one another. This can be avoided by using the zone schemes presented earlier. Figure 5.19b simply gives the area of the Fermi surface for the first Brillouin zone, while Figure 5.19c shows the Fermi surface in the second zone as a diagram of reduced zones (that is to say brought back as in the first zone). So that Figure 5.19c is not overloaded, only the contribution from the 2_a part of the second Brillouin zone is shown (as a dotted surface). This figure can be compared with that in Figure 5.17.

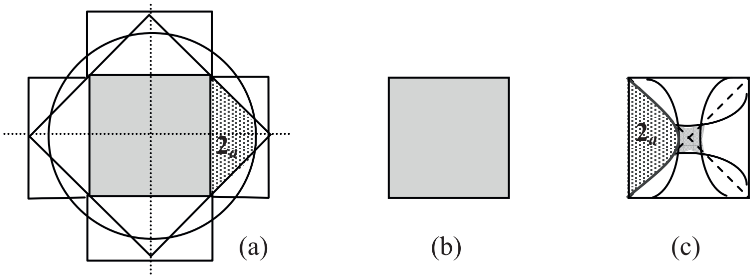


Figure 5.19. *(a) Circular Fermi surface for an arbitrary concentration of electrons; (b) part of the Fermi surface situated in the first zone; and (c) part of the Fermi surface situated in the second zone brought to the first zone with a contribution from 2a, shown as the dotted area. The central areas are the only ones not occupied by electrons.*

5.9.3. *Evolution of semi-free electron Fermi surfaces*

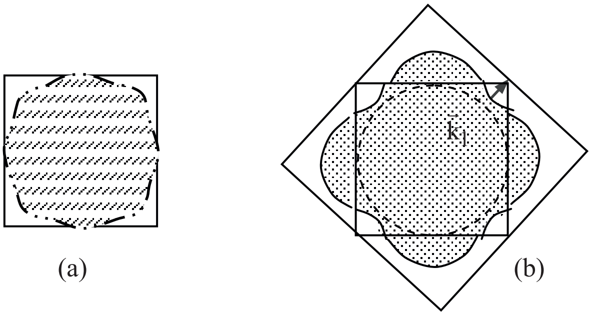
The filling-up of electronic levels is accomplished from the bottom level up (the origin of the reciprocal lattice), and also in the corresponding equi-energy circular levels (or spheres in 3D) of which the radius (k) increases along with the filling. From a circle (or a sphere) of radius greater than $2\frac{\pi}{a}$ (side of the square representing the first Brillouin zone), that is to say from the limit of the Brillouin zone in the directions k_x and k_y (see Figure 5.13) the second zone will start filling before the first Brillouin zone is completely filled. There is an overlap between zones, as long as the potential barrier between the Brillouin zone is small.

It should be said that at this level there is a distortion caused by Brillouin zones being defined from a periodic potential. So while the sphere (or circle in 2D) is drawn based on a hypothesis of free electrons (using the relation $E = \frac{\hbar^2}{2m}(k_x^2 + k_y^2 + k_z^2)$ for free electrons, the electrons are actually semi-free.

So as to have a reasoning closer to reality, we need to have an idea of the form of the Fermi surfaces in an approximation of semi-free electrons. Qualitatively, we can state that (and this applies to a base of atoms liberating two electrons so as to fill the first zone, as is the case for bivalent atoms with one atom per base):

– if the barrier is sufficiently high, the electrons remain in the first zone (as for example, in diamond) and as shown in Figure 5.20a. First they are placed around the center and then fill the first zone up to the point where the circle (sphere) just touches the boundaries. Then they fill up the areas towards the corners, as they do not have enough energy to break the barrier to the second zone;

– if the barrier is not too high, electrons in the second zone are less energetic than those in the first, and now Figure 5.20b gives a better representation of the minimum energy.



(a) (b)

Figure 5.20. *Deformation of at the limit of the Fermi surface when there is: (a) a high barrier; and (b) a small barrier*

Approximate constructions of semi-free electron Fermi surfaces are based on the following two facts:

i) *The Fermi surface meets the limits of the Brillouin zone at a right angle*

In effect, the velocity of the group associated with the electron wave packet is written as:

$$v_g = \frac{d\omega}{dk} \overset{\omega=\frac{E}{\hbar}}{=} \frac{1}{\hbar}\frac{dE}{dk},$$

so that vectorially $\vec{v}_g = \frac{1}{\hbar}\overrightarrow{\text{grad}_k} E$. The vector $\overrightarrow{\text{grad}_k} E$ being normal to equi-energy lines traced in k space (reciprocal space) means that the equi-energy lines are normal to $\vec{v}_g$.

Looking at a Brillouin zone limit, and working in terms of H as shown in Figure 5.21, the wave vector y is written as: $\vec{k} = [\vec{k}_x]_{k=\frac{\pi}{a}} + \vec{k}_y$.

The velocity (y) is given in the form: $\vec{v} = \vec{v}_x + \vec{v}_y$, and is collinear with $\vec{k}$ as $\hbar\vec{k} = m^*\vec{v}$ (crystalline moment).

As $\left(\frac{dE}{dk}\right)_{k_x=\pi/a} = 0$, $v_x = 0$ and $\vec{v} = \vec{v}_y \parallel \vec{k}_y$, the velocity is directed with respect to $\vec{k}_y$ and the equi-energy lines are normal to the zone limit (limit directed with respect to $\vec{k}_y$).

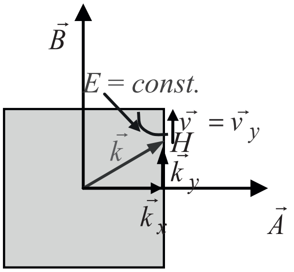


Figure 5.21. *Electron velocity in the band limits*

ii) *The crystal potential has the effect of bending the Fermi surface close to the energy extremes*

In effect, within a vector $\vec{k}$ of the energy extreme (defined in the reciprocal space by, for example, $\vec{k}_n \overset{n=1}{=} \vec{k}_1$), the energy that is developed (following MacLaurin) can be written using:

$$E(k) = E(k_1) \pm \frac{\hbar^2 (k - k_1)^2}{2|m^*|}.$$

The equi-energy surfaces are spheres in #D or circles in 2D with centers at $\vec{k}_1$ and radii defined by $\left(k - k_1\right) = \frac{\sqrt{2|m^*|(E - E_1)}}{\hbar}$. This explains the shape of the equi-energy spheres at the limit of the zone.

5.9.4. *Relation between Fermi surfaces and dispersion curves*

As already discussed, we should consider all possible directions in the reciprocal lattice because an energy that is forbidden in one direction (k_x in $k_x = \pi/a$) may be permitted in another (k_{xy} again with $k_x = \pi/a$). This configuration is described in

Figure 5.22 for a 2D Brillouin zone of a material with a square lattice (equivalent to a simple cubic lattice in 3D).

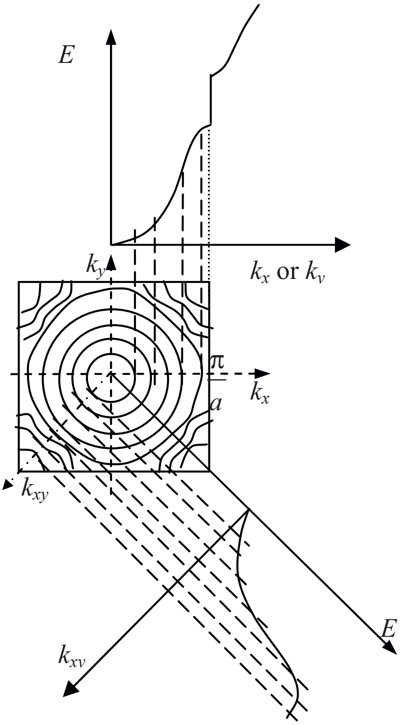


Figure 5.22. *Relation between the curves of* $E = f(k)$ *and equi-energy*

The circles shown in the zone are the geometric points for an energy E_N where $N = 1, 2, 3$, etc. The circles are pushed one against another as N increases. When we reach the zone limit, the circle is deformed and at the very limit a discontinuity of the equi-energy curves occurs. While energies are forbidden in the k_x and k_y directions, they are still allowed in the k_{xy} direction, up to the point where the whole zone is filled if the gap between the two zones is large enough.

In general terms, we cannot obtain any quantitative results without performing some calculations but qualitatively we can see that the equi-energy surface of the second Brillouin zone (Figure 5.19c for free electrons) will develop into a form shown in Figure 5.23.

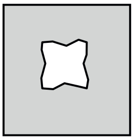


Figure 5.23. *Fermi surface (in gray) in the second Brillouin zone (scheme f reduced zones)*

5.10. Conclusion. Filling Fermi surfaces and the distinctions between insulators, semiconductors and metals

5.10.1. *Distribution of semi-free electrons at absolute zero*

Generally, at absolute zero, the electrons occupy the minimum energy cells close to the origin of the reciprocal lattice. These cells are distributed inside the Fermi surface.

If the element making up the solid liberates few electrons per atom, for example crystalline monovalent elements in a centered cubic system (alkali metals) or in a face centered cubic lattice (e.g. *Cu*, *Ag*, *Au* as detailed in Chapter 6), the number of occupied cells is much lower than the number of cells contained in the first Brillouin zone. The equi-energy surface (Fermi-surface) is very close to that of a sphere (see Figure 5.24 below for a square lattice) and the representation of the free electrons scheme is very close to reality.

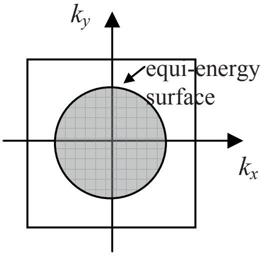


Figure 5.24. *Representation of a square lattice*

If the atoms of the element making up the solid liberate two electrons or more (in general an even number), then the first (or the next) Brillouin zone can be practically totally filled. However, there are two particular situations that can arise:

– if the gap between the first and the second Brillouin zone (or more generally between adjacent Brillouin zones) is large, then the Brillouin zone under consideration can end up being totally filled, as in Figure 5.25 below;

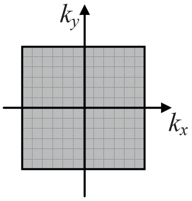


Figure 5.25. *A totally filled Brillouin zone*

– if the gap is small, or even better the bands overlap, causing the gap to disappear, then the second zone starts to fill up before the first zone under consideration is completely full.

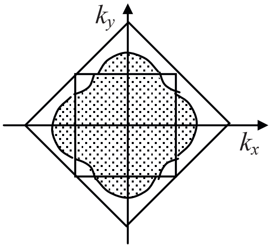


Figure 5.26. *A partially filled Brillouin zone*

5.10.2. *Consequences for metals, insulators/semiconductors and semi-metals*

If the last occupied zone (band) is only partially filled (generally half-way) there are numerous empty cells that remain available to transport electrons under the effect of an external perturbation (for alkali metals and monovalent noble metals see the example of copper given in Chapter 6).

If the last occupied zone (band) is completely full, and if the gap with the following zone (band) is considerable (greater than 4 or 5 eV), the electrons cannot leave the band under consideration and the material is an insulator. If the gap is relatively small (typically less than 3 or 4 eV), the insulating material at absolute zero becomes a semiconductor at ambient temperature as there are a few electrons that can pass with thermal agitation to the following zone (band) to leave holes behind (for example see germanium or silicon detailed in Chapter 6). The presence of a reasonably high gap (in the order of 5 eV such as in carbon based diamond) can nevertheless be used to furnish a material with semiconductor properties that can resist intense "flashes", such as nuclear explosions, and can be used in devices of military importance.

If the last band can be totally filled but there is an overlap of bands, we end up with a semi-metal (for example, alkaline-earth metals have two valent electrons per primitive cell, which would be insulators except that the energy bands overlap and hence they have to a limited extent some metallic properties).

5.11. Problems

5.11.1. *Problem 1: simple square lattice*

Show that for a simple square lattice (2D) that kinetic energy of a free electron in a corner of the first Brillouin zone is twice as high as that of an electron situated in the middle of one side of the zone. What is the result when the lattice is cubic (3D)?

Answer

For a free electron, the kinetic energy is written: $E = \frac{\hbar^2 k^2}{2m}$, and we find in 2D that:

– in the middle of one side of the zone: $k^2 = k_x^2 = \left(\frac{\pi}{a}\right)^2$, and $E_{\text{side}} = \frac{\hbar^2}{2m}\left(\frac{\pi}{a}\right)^2$;

– and at a corner of the zone: $k^2 = k_x^2 + k_y^2 = 2\left(\frac{\pi}{a}\right)^2$ and $E_{\text{side}} = \frac{\hbar^2}{2m}2\left(\frac{\pi}{a}\right)^2$.

We thus find that:

$$\frac{E_{\text{corner}}}{E_{\text{side}}} = 2.$$

In a 3D medium, the value of E_{side} is unchanged while:

$$E_{\text{corner}} = \frac{\hbar^2}{2m}(k_x^2 + k_y^2 + k_z^2) = \frac{\hbar^2}{2m}3\left(\frac{\pi}{a}\right)^2,$$

from which in 3D: $\frac{E_{\text{corner}}}{E_{\text{side}}} = 3.$

5.11.2. *Problem 2: linear chain and a square lattice*

1) This question concerns a linear chain made up of n identical atoms that are spaced an equal distance apart from one another so that the length of the chain L is given by $L \approx na$.

a) What is the dimension of a cell in the reciprocal space?

b) How many cells can we place in each energy band?

c) Show that a 1D body is an electrical insulator (or semiconductor) at $T = 0K$ if it carries an even number of valence electrons per atom.

2) This question carries on from the question concerning a square lattice (of side $L \approx na$) that contains in all N atoms ($N = n^2$) and each primitive cell has a side of length a.

a) Describe the direct lattice and then the reciprocal lattice.

b) Trace the first two Brillouin zones. Indicate the maximum number of electrons that can be placed in the first and the second zone. Conclude.

3) The atoms in this question are bivalent.

a) Sketch the scheme of the energy bands for semi-free electrons: $E = f(k_{1,0})$ and $E = \mathrm{f}(k_{1,1})$ (where k_{ij} designates the wavenumber in the direction $[i, j]$).

b) Point X designates the middle of the side of the square that is the first zone. Point M is at the summit of the same square. The modulus of the wave vectors have their ends at X and at M and are denoted by k_X and k_M. When $k = k_X$, the corresponding energies are denoted $E(X_C)$ for the bottom of the conduction band and $E(X_V)$ for the summit of the valence band. In addition $E(X_C) - E(X_V) = w_{1,0}$ and is the energy gap in the direction [1, 0].

Similarly, when $k = k_M$, the energy is denoted $E(M_C)$ for the bottom of the conduction band and $E(M_V)$ for the summit of the valence band, with $E(M_C) - E(M_V) = w_{1,1}$ being the energy gap in the direction [1, 1].

What relations are there between $E(M_V)$ and $E(X_C)$ when the material that is made up of bivalent atoms is at $T = 0K$ an insulator or a conductor?

c) Within the last hypothesis, detail the appearance of the Fermi and first Brillouin zone surfaces.

4) Numerical analysis:

$w_{1,0} = 4$ eV, $w_{1,1} = 2$ eV,

$$\frac{\hbar^2}{2m}\frac{\pi^2}{a^2} \simeq 4.2 \text{ eV (when } a \approx 0.3 \text{ nm).}$$

Is the material a conductor or insulator at absolute zero?

Answers

1)

a) For a direct space of dimension L, the dimension of one cell in reciprocal space is given by $\Delta k = \frac{2\pi}{L}$, where $L = na$, with n making p the number of atoms that go into a linear chain.

b) For a periodic system of period a, an energy band is localized between $k = -\frac{\pi}{a}$ and $k = \frac{\pi}{a}$. Its width is thus given by $\frac{2\pi}{a}$, and inside it can be placed $\frac{\frac{2\pi}{a}}{\frac{2\pi}{L}} = \frac{L}{a} = n$ cells.

c) With the possibility of placing two electrons per cell, we can thus place $2n$ electrons per band (each band being the same size, as in $\frac{2\pi}{a}$) in all. If the atoms each liberate an even number (N_p) of valence electrons, then the number of bands that the nN_p of electrons that will be able to fill at absolute zero (temperature at which the Fermi function is equal to one when $E < E_V < E_F$) will be:

– 1 band if $N_p = 2$,

– 2 band if $N_p = 4$,

– 3 band if $N_p = 6$,

– etc.

with, in all cases, a last band that is completely full. The material is thus an insulator if there is a large gap and a semiconductor if the gap is small (i.e. less than 3 eV).

2)

a) The problem is now 2D with a square direct lattice of side $L = na$ and $N = n^2$ atoms in all.

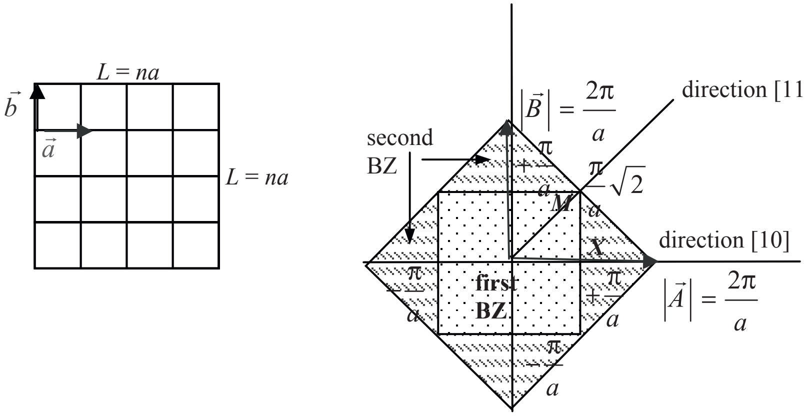


Figure 5.27. *Sketches for Answer 2: square direct lattice (left) and the base vectors on the reciprocal lattice (right)*

The base vectors of the reciprocal lattice are such that $|\vec{A}| = |\vec{B}| = \frac{2\pi}{a}$, where $\vec{A} \cdot \vec{b} = 0$, so that $\vec{A} \perp \vec{b}$ and similarly $\vec{B} \perp \vec{a}$.

b) The first two zones shown above are of the same surface (as detailed in many university courses on how to use reduced zones) and equal to $\left(\frac{2\pi}{a}\right)^2$. The 2D cells in reciprocal space are of a size given by $\Delta k_x \cdot \Delta k_y = \frac{4\pi^2}{L^2}$, and therefore in each zone we can place $\left(\frac{2\pi}{a}\right)^2 \Big/ \left(\frac{2\pi}{L}\right)^2 = \frac{L^2}{a^2} = n^2 = N$ cells, which is equal to the number of atoms in the system. With two electrons per cell, we can put in $2N$ electrons per zone. If the N atoms have a valency of two, they will liberate $2N$ electrons that can only just fully fill the first zone. This complete filling will only occur if the barrier between the two zones is sufficiently high to push electrons to the corners at the end, and will result in insulating (or semiconducting) behavior. If the atoms have an odd number valency (monovalent), the last band to be occupied will be half-full in respect of equi-energy circles with a maximum radius of $k_f = \frac{\sqrt{2m^* E_F}}{\hbar}$. If $k_f < \frac{\pi}{a}$, then the circles do not reach the second Brillouin zone, the bands will not overlap and a metal is formed. If $k_f > \frac{\pi}{a}$, there is a risk of two bands overlapping with the ensuing formation of semi-metallic behavior.

3) The atoms are bivalent, and when the extremity of a wave vector touches the zone limit, the energy levels of the different bands need to be compared.

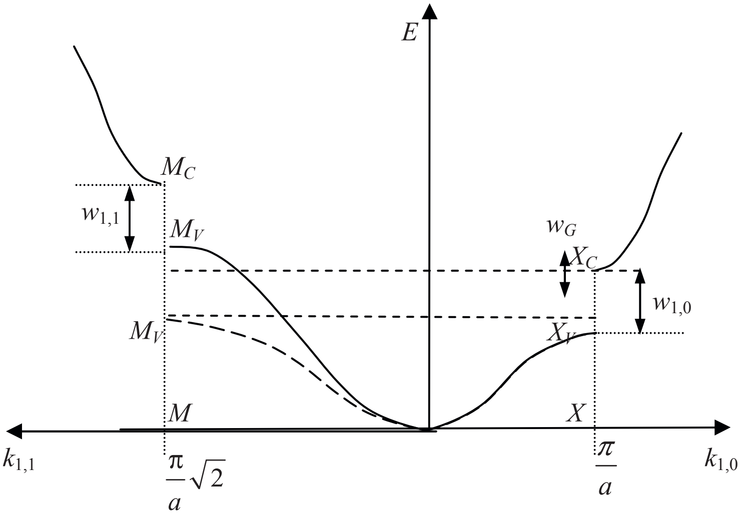


Figure 5.28. *Diagram of the band scheme with respect to the directions [1,0] and [1,1]*

a) Schematically, and from the perspective of the bands, the representation above shows the two principal directions [1, 0] and [1, 1]:

– If $E(M_V) < E(X_C)$, the first band and hence the first Brillouin zone is completely filled with up to $2N$ electrons. Being completely filled it is thought of as an insulator or semiconductor with a gap given by: $E(X_C) - E(M_V) = w_G$.

– If $E(M_V) > E(X_C)$, the second band starts to fill before the first band is completely full. The two bands are incompletely filled and have overlapping energy bands – the materials is semi-metallic (no resultant gap). Within the free electron theory, the Fermi surface takes on the shape shown below as a dashed line circle. For these semi-free electrons, the circle becomes deformed as shown in section 5.9.3.

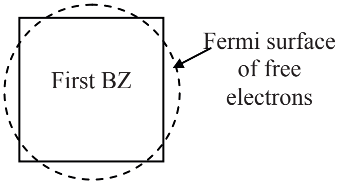


Figure 5.29. *Incomplete bands with overlapping energy levels (as when* $E(M_V) > E(X_C)$*)*

These ideas can be extended to 3D materials and can explain how *Mg*, *Be* and *Pb* are metallic, while *Si*, *Ge* and *Se* are semiconductors.

Given the relationship that exists at the band limit, we can quantitatively write that for the preceding two inequalities in a square lattice:

$$k_X = \frac{\pi}{a}$$

$$k_M = \frac{\pi}{a}\sqrt{2}.$$

Knowing $w_{1,1}$ and $w_{1,0}$, and hence $\frac{\hbar^2}{2m}\frac{\pi^2}{a^2} \simeq 4.2\,\text{eV}$, we can discover whether or not the bivalent material is an insulator or a semi-metal.

With $w_{1.0} = 4$ eV and $w_{1,1} = 2$ eV we have:

$$E(M_V) = \frac{\hbar^2}{2m}k_{11}^2 - \frac{w_{11}}{2} = \frac{\hbar^2}{2m}\frac{2\pi^2}{a^2} - \frac{w_{11}}{2}$$

$$= 4.2 \times 2 - \frac{2}{2} = 7.4 \text{ eV}$$

$$E(X_C) = \frac{\hbar^2}{2m}k_{10}^2 + \frac{w_{10}}{2} = \frac{\hbar^2}{2m}\frac{\pi^2}{a^2} + \frac{w_{10}}{2}$$

$$= 4.2 + \frac{4}{2} = 6.2 \text{ eV}.$$

Thus $E(M_V) > E(X_C)$, which implies a semi-metallic character.

NOTE.– In the numerical calculation, there is:

$$E_{\pi/a} = \frac{\hbar^2}{2m}\frac{\pi^2}{a^2} = \frac{h^2}{8m}\frac{1}{a^2} \overset{a=0.3\text{nm}}{=} \frac{\left(6.62x10^{-34}\right)^2}{8\times 0.9\times 10^{-30}}\frac{1}{9\times 10^{-20}} = 6.76\times 10^{-19}\,\text{J},$$

or rather:

$$E_{\pi/a} = \frac{6.76\times 10^{-19}}{1.6\times 10^{-19}}\text{eV} = 4.227 \text{ eV}.$$

5.11.3. *Problem 3: rectangular lattice*

This problem concerns a 2D material in a rectangular lattice that has the following parameters: the primitive cell is such that $a = 2$ Å and $b = 4$ Å; the lengths of the real crystal are $L_1 = N_1 a$ and $L_2 = N_2 b$; and the number of nodes (where there are placed A-type atoms) is given by $N \approx N_1 N_2$. We assume that the base is defined by two atoms: atom A placed at (0, 0) and atom B placed at (1/4, 1/4). The outer layer electrons dispersion relationship (valence electrons) follows that of the semi-free electrons theory. That means for a given direction $[m, n]$ the curve $E = f(k)$ follows that of free electrons (of energy E) except when in the vicinity of discontinuities, where the energy for the first discontinuity can be written as:

$E_{m,n} = E_{m,n} \pm w_{m,n}/2$ (discontinuity amplitude is $w_{m,n}$ where in this case $w_{m,n} > 0$).

In terms of notation:

$$E_V[m, n] = E^{-}_{m,n} = E_{m,n} - w_{m,n}/2 \text{ and } E_C[m, n] = E^{+}_{m,n} = E_{m,n} + w_{m,n}/2.$$

And we have $\hbar^2/2m = 3.82$ eV Å^2.

1) Show:

a) the direct lattice with base vectors that are written as $\vec{a} = a\vec{e}_x$ and $\vec{b} = b\vec{e}_y$;

b) the reciprocal lattice and the base vectors that we will determine the first two Brillouin zones and their geometric forms.

2) Give the expression for the energy at zero order in k space at the first point of discontinuity that appears with the first-order approximation (semi-free electrons) in:

a) the direction [1, 0];

b) the direction [0, 1];

c) the direction [1, 1].

In the figure also sketch the dispersion curves $[E = f(k_{m,n})]$ for semi-free electrons with various energy levels, namely: $w_{1.0} = 1.5$ eV; $w_{0.1} = 2$ eV; and $w_{1.1} = 1$ eV.

3) This question studies the filling of Brillouin zones at absolute zero. Each AB atomic basis set liberates two electrons that are semi-free.

a) Indicate the number of electrons that the first Brillouin zone can accommodate.

b) What inequality exists between the parameters $w_{m,n}$ and a and b of the direct lattice cell so that a crystal is an insulator?

4) Applying the result from above.

a) The values that are given are: $w_{1,0}$ = 1.5 eV; $w_{0,1}$ = 2 eV; $w_{1,1}$ = 1 eV; $\hbar^2/2m$ = 3.82 eV Å^2. Determine whether or not the crystal is an insulator or a conductor at absolute zero.

b) Now the material is a conductor at absolute zero. Schematically show the form of the equi-energy curve E_F (corresponding to the Fermi level for which the significance of absolute zero should be noted).

5) From now on *B* is placed at (1/2, 1/2) and the atoms *A* and *B* are chemically identical. The direct lattice can now be thought of as part of a centered rectangle.

a) Show the geometric form of the direct lattice with its new primitive cell (along with the two fundamental vectors and their components).

b) Give the base vectors of a traced reciprocal lattice and give its structure.

c) Trace the first Brillouin zone and show its geometric form.

d) We will assume that the *A* atoms are monovalent. From the determination of the Fermi circle line, show if the material is a conductor or an insulator.

e) The *A* atoms are now supposed to be bivalent. Indicate the new position of the Fermi circle. From this can it be deduced, using the given data, whether the material is an insulator or a conductor?

Answers

1)

a)

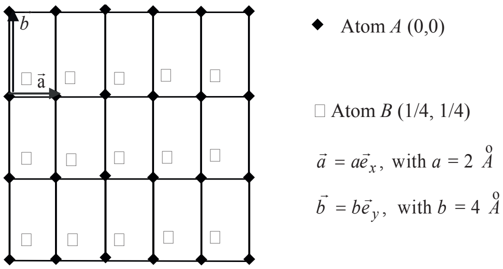
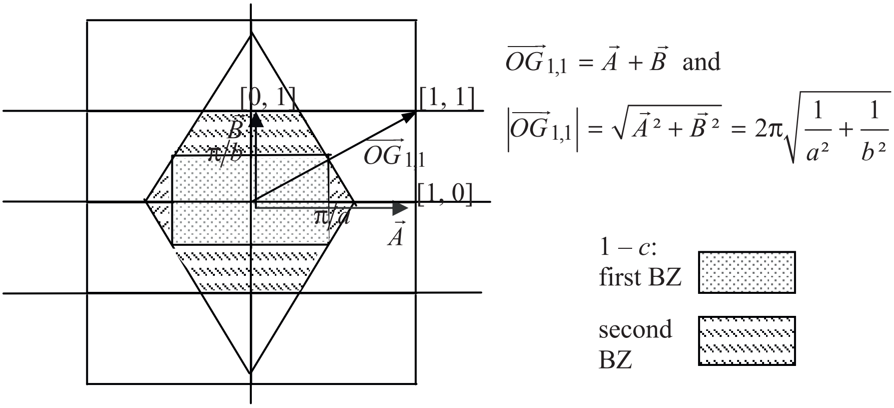


Figure 5.30. *Direct lattice with base vectors*

b) The base vectors of the reciprocal lattice are such that $\left|\vec{A}\right| = \dfrac{2\pi}{a}$, with $\vec{A} \cdot \vec{b} = 0$, so that $\vec{A} \perp \vec{b}$; and similarly, $\left|\vec{B}\right| = \dfrac{2\pi}{b}$, with $\vec{B} \cdot \vec{a} = 0$, we have $\vec{B} \perp \vec{b}$.



$\overrightarrow{OG}_{1,1} = \vec{A} + \vec{B}$ and

$$\left|\overrightarrow{OG}_{1,1}\right| = \sqrt{\vec{A}^2 + \vec{B}^2} = 2\pi\sqrt{\frac{1}{a^2} + \frac{1}{b^2}}$$

Figure 5.31. *Reciprocal lattice with base vectors*

In the figure it can be seen that the two upper and lower triangles are empty. They do not belong to the second Brillouin zone but to those at higher numbers.

2) The answer can be given as:

a) In the [1, 0] direction (which is that of $\vec{A}$): $k = k_{1,0} = \frac{\pi}{a}$; and we thus have $a = 2$ Å

$$E^0_{1,0}\left(k = \frac{\pi}{a} = 1.57\right) = \frac{\hbar^2}{2m}\left(\frac{\pi}{a}\right)^2 = 3.82\frac{\pi^2}{4} = 9.43 \text{ eV}$$

b) In the [0, 1] direction (which is that of $\vec{B}$): $k = k_{0,1} = \frac{\pi}{b}$; and we thus have $b = 4$ Å

$$E^0_{1,0}\left(k = \frac{\pi}{b} = 0.79\right) = \frac{\hbar^2}{2m}\left(\frac{\pi}{b}\right)^2 = 3.82\frac{\pi^2}{16} = 2.36 \text{ eV}$$

c) In the [1, 1] direction (which is that of $\vec{G}_{1,1}$): $k = k_{1,0} = \frac{\pi}{a}$; and

$$E^0_{1,1}\left(\left|\vec{k}\right| = \pi\sqrt{1/a^2 + 1/b^2} = 1.76\right)$$

$$= \frac{\hbar^2}{2m}\left(\left[\frac{\pi}{a}\right]^2 + \left[\frac{\pi}{b}\right]^2\right) = 3.82\left(\frac{5\pi^2}{16}\right) = 11.78 \text{ eV.}$$

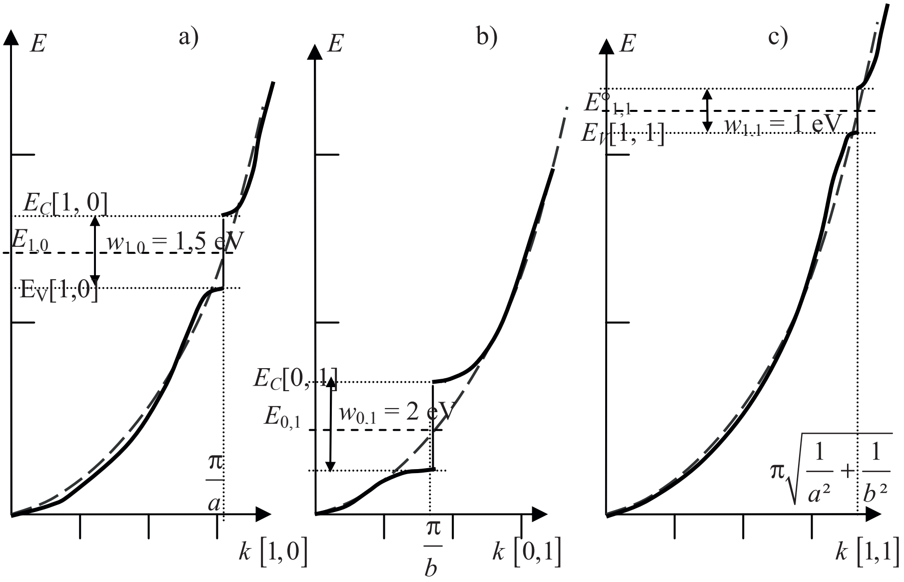


Figure 5.32. *Band structures in the directions: a)* [1,0]*;b)* [0,1]*; c)* [1,1]

3) Filling Brillouin zones:

a) Each atomic base liberates two electrons and in the direct lattice we have on average one base per rectangular cell (four bases are attached to each cell, and each base is shared between four cells so there is on average one atom base per cell). If S is defined as the surface of the direct lattice, then $S = L_1L_2 = N_1N_2ab$. With each cell having the surface ab, we find $S/ab = N_1N_2 = N$ which is the number of atoms (or nodes) in the direct lattice.

In the reciprocal lattice, the surface of the first Brillouin zone is given by $S^* = \left(\frac{2\pi}{a}\right)\left(\frac{2\pi}{b}\right) = \frac{4\pi^2}{ab}$. The surface of a primitive cell is such that $\Delta k_x \Delta k_y = \left(\frac{2\pi}{L_1}\right)\left(\frac{2\pi}{L_2}\right) = \frac{4\pi^2}{N_1 N_2 ab}$, so that in the first Brillouin zone we can place $\frac{4\pi^2}{ab}\Big/\frac{4\pi^2}{N_1 N_2 ab} = N_1 N_2 = N$ cells. Being able to place two electrons per cell, the first Brillouin zone is totally full with $2N$ electrons. As a consequence, the first Brillouin zone may receive $2N$ electrons freed by the base of N atoms.

b) In order to determine if the material is an insulator/semiconductor or semi-metallic, we need to know if the bands are overlapping or not. This means finding out whether or not the band filling in the direction [0, 1] (corresponding to the lowest energies) of the second band starts before filling of the first band in other directions is complete.

So that there is no overlapping whatsoever, the lowest level of the conduction band should not be filled before the highest level of the valence band is complete. For this to happen, there should simply be no overlap between the directions [0, 1] and [1, 1] that correspond, respectively, to the lowest and highest energy directions.

In order to have the insulating (or semiconducting) state, we therefore need to find:

$$E_V[1, 1] < E_C[0, 1],$$

so that:

$$E^-_{1,1} = E_{1,1} - w_{1,1}/2 < E^+_{0,1} = E_{0,1} + w_{0,1}/2.$$

This condition can also be written as:

$$w_{0,1} + w_{1,1} > 2(E_{1,1} - E_{0,1}) = \frac{2\hbar^2}{2m}\left[\pi^2\left(\frac{1}{a^2} + \frac{1}{b^2}\right) - \pi^2\left(\frac{1}{b^2}\right)\right],$$

so that:

$$w_{0,1} + w_{1,1} > \frac{\hbar^2\pi^2}{ma^2}.$$

4) Numeric application:

a) With $w_{0,1} = 2$ eV and $w_{1,1} = 1$ eV, we have: $w_{0,1} + w_{1,1} = 3$ eV.

With $a = 2$ Å and $\hbar^2/2m = 3.82$ eV Å^2, we can deduce:

$$\frac{\hbar^2\pi^2}{ma^2} = 18.85 \text{ eV, from which } w_{0,1} + w_{1,1} = 3 \text{ eV} < \frac{\hbar^2\pi^2}{ma^2} = 18.85 \text{ eV}.$$

The material is therefore semi-metallic.

b) If the material is a conductor, we have: $E^{\circ}_{0,1} + w_{0,1}/2 > E_{1,1} - w_{1,1}/2$. At 0 K, the Fermi level (E_F) penetrates the second band along the direction [0, 1] as E_F must be higher than $E_C[0, 1]$. Nevertheless, it will be at the interior of the first band in the [1, 1] direction as it is in this direction that E_F must be less than $E_V[1, 1]$. This can also be compared against the figure in problem 2, which describes the characteristics for a semi-metallic state.

In the direction [1, 1] the Fermi ray $k_F[1, 1]$ increases continually while in the direction [0, 1], $k_F[0, 1]$ has to go through a step before starting to increase again. On taking into account the deformation of the Fermi circles near the angles, we find that the Fermi surface takes on the form shown below.

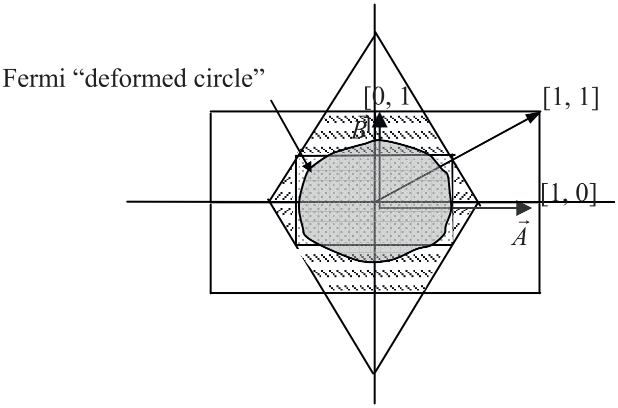


Figure 5.33. *Truncated Brillouin zone*

5) The B atoms are now placed at (1/2, 1/2) and are the same as the A atoms. They are pictured below:

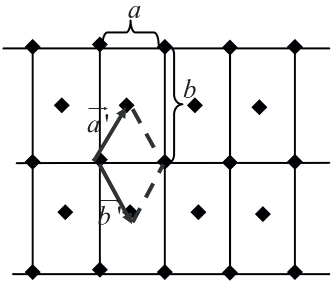


Figure 5.34. *Direct latice with B atom localized in (1/2, 1/2)*

a) The direct lattice is thus a centered rectangle. The primitive cell is rhombus shaped, and a base vector for the vectors such that:

$$\vec{a}'\begin{cases} a/2 \\ b/2 \end{cases} \text{ and } \vec{b}'\begin{cases} a/2 \\ -b/2 \end{cases}.$$

The surface of the elementary cell is:

$$s' = 4 \cdot \frac{1}{2} \cdot \left(\frac{a}{2}\frac{b}{2}\right) = \frac{a \cdot b}{2},$$

and the number of elementary cells that we can place in a sample of the surface is given by $S = L_1L_2 = N_1N_2\ ab = Nab$ and is equal to $\frac{S}{s'} = 2N$. As each rhombic primitive cell contains on average one atom (four atoms shared between four adjacent cells), a sample of the surface S contains $2N$ atoms in all.

NOTE.– If instead of considering the preceding primitive cell, we look at a centered rectangular cell, the surface is now given by $s = ab$. In the sample given by the surface $S = L_1L_2 = N_1N_2\ ab = Nab$, we can place $S/s = N$ rectangular cells. As this cell contains on average two atoms per cell (one atom at the center of the cell and on average one atom at the top of the cell, with four atoms at four summits being shared between four adjacent cells), we can place a total of $2N$ atoms in the sample surface S, i.e. a result identical to the preceding case.

b) The base vectors of the reciprocal lattice, $\overrightarrow{A'^*}\begin{cases} X \\ Y \end{cases}$ and $B'^*\begin{cases} V \\ W \end{cases}$ are such that they should agree with:

$$\overrightarrow{A'^*} \cdot \vec{a}' = X\frac{a}{2} + Y\frac{b}{2} = 2\pi \quad (1) \qquad \overrightarrow{B'^*} \cdot \vec{b}' = V\frac{a}{2} - W\frac{b}{2} = 2\pi \quad (2)$$

$$\overrightarrow{A'^*} \cdot \vec{b}' = X\frac{a}{2} - Y\frac{b}{2} = 0 \quad (3) \qquad \overrightarrow{B'^*} \cdot \vec{a}' = V\frac{a}{2} + W\frac{b}{2} = 0 \quad (4)$$

We thus have a system based on four equations with four unknowns (X, Y, V, W). So:

$$\left.\begin{array}{l}(1)+(3)\Rightarrow X\,a=2\,\pi\Rightarrow X=2\pi/a\\(1)-(3)\Rightarrow Y\,b=2\,\pi\Rightarrow Y=2\pi/b\\(2)+(4)\Rightarrow V\,a=2\,\pi\Rightarrow V=2\pi/a\\(2)-(4)\Rightarrow -W\,b=2\,\pi\Rightarrow W=-2\pi/b\end{array}\right\}\Rightarrow \overrightarrow{A'^*}\begin{cases}2\pi/a\\2\pi/b\end{cases}\text{ and }\overrightarrow{B'^*}\begin{cases}2\pi/a\\-2\pi/b\end{cases}$$

As a consequence, the elementary cell of the reciprocal lattice has the same conformation as the reciprocal lattice cell. The reciprocal lattice is also a centered rectangle. Mathematically, we in effect need only to equate the components of $\overrightarrow{A'^*}$ and $\overrightarrow{B'^*}$, as in $2\pi/a = c/2$ and $2\pi/b = d/2$, to see that the fundamental vectors $\overrightarrow{A'^*}$ and $\overrightarrow{B'^*}$ have the same components as the fundamental vectors $\overrightarrow{a'}$ and $\overrightarrow{b'}$ from the direct lattice when:

$$\overrightarrow{A'^*}\begin{cases}c/2\\d/2\end{cases}\quad\text{and}\quad\overrightarrow{B'^*}\begin{cases}c/2\\-d/2\end{cases}.$$

c) The first Brillouin zone takes on a rhombic shape – shown as a gray surface in the figure below.

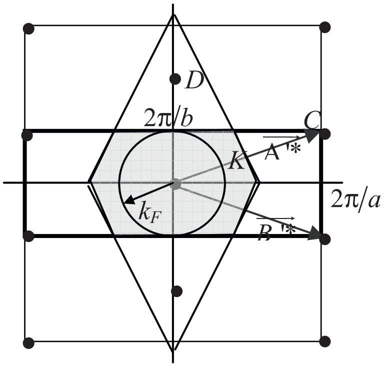


Figure 5.35. *Rhombic-shaped first Billouin zone (gray area)*

We can state that:

$$\mathrm{OC} = 2\pi\sqrt{\frac{1}{\mathrm{a}^2}+\frac{1}{\mathrm{b}^2}}\ ;$$

and its mediator delimits (in the direct lattice of the centered rectangle) the first Brillouin zone, and is identical to that which delimits the second Brillouin zone of the direct lattice of the preceding example based on a simple rectangle. The same is true for $OD = 2.2\pi/b$ and its mediator.

We can immediately conclude that the surface of the first Brillouin zone is the centered rectangular lattice is equal to the sum of the surface of the first Brillouin zone and second Brillouin zone of the preceding simple rectangle lattice.

d) As seen in problem 5a where there were $2N$ electrons to place, in the direct lattice if the A type atoms are monovalent, we have $2N$ atoms to place that take up the reciprocal space of N cells on a surface given by: $\left(\frac{2\pi}{L_1}\right)\left(\frac{2\pi}{L_2}\right) = \frac{4\pi^2}{Nab}$.

These cells should be spread throughout the Brillouin zones with the first zone being filled first, it having dimensions given by:

$$OL = \frac{2\pi}{b} = \frac{2\pi}{4} = \frac{\pi}{4} = 1.57\ \text{Å}^{-1}$$

$$OK = \frac{1}{2}\sqrt{\left(\frac{2\pi}{a}\right)^2 + \left(\frac{2\pi}{b}\right)^2} = \pi\sqrt{\frac{1}{a^2}+\frac{1}{b^2}} = 1.76\ \text{Å}^{-1}$$

The line detailed by k_F of the Fermi circle, at the interior of which are placed all the $2N$ electrons of the N cells if the line is inside the Brillouin zone is such that $\pi k_F^2 = N \cdot (\text{Surface of a cell}) = \frac{4\pi^2}{ab}$. From this can be deduced that $k_F = 2\sqrt{\frac{\pi}{ab}} \overset{\substack{a=2\\b=4}}{=} 1.26\ \text{Å}^{-1}$.

The Fermi circle thus appears smaller than the smallest dimension (here along OL) of the first Brillouin zone so that all the electrons can be placed inside the Fermi circle while maintaining empty cells in all directions. The material is therefore a conductor.

e) If the *A* atoms are divalent, we have 4*N* electrons to place into 2*N* cells, and the Fermi circle must therefore follow: $\pi k_F^2 = 2N \cdot \left(\frac{4\pi^2}{Nab}\right) = \frac{8\pi^2}{ab}$, so that $k_F = 2\sqrt{\frac{2\pi}{ab}} \overset{\substack{a=2\\b=4}}{=} 1.78$ Å^{-1}. This time, $k_F > OL$ and even $k_F > OK$, so the Fermi circle reaches the second zone and what remains to be known is what happens at the points *L* and *K*:

– either there is little or no potential barrier and the second zone starts to be populated before the first is completely full. The material is thus a semi-metal;

– or there is a high barrier that displaces electrons into the corners of the first zone and we thus need to know if we can just fill the first zone of the rectangular center, in which case the material would be an insulator.

Taking the representation for the Brillouin zone from problem 1b, it is possible to see that with the simple rectangular structure we can use the first zone's scheme of reduced zones for the second zone, which has the same surface. The figure below shows how part (1) of the second zone can be placed into the first. The same can be done with other parts of the second zone, as for example the surfaces of small triangles t_1 and t'_1 are identical. The result is that we can place, in all, 2 × 2*N* electrons, or rather, 4*N* electrons in the first two Brillouin zones of the centered rectangular structure. This is just the number of electrons liberated by the bivalent *A* atoms. In effect, the material is an insulator.

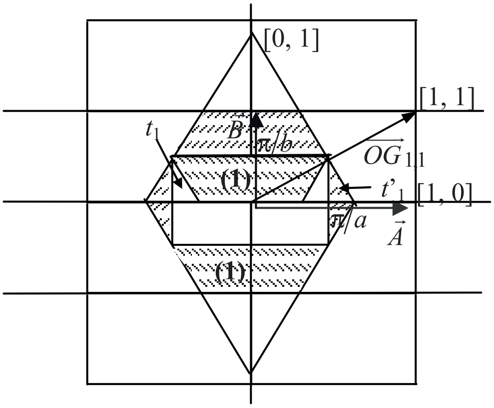


Figure 5.36. *Diagram of the reduced zones showing that first and second zones are of the same area*

Chapter 6

Electronic Properties of Copper and Silicon

6.1. Introduction

This chapter is mostly concerned with the face centered cubic (fcc) structure that is taken up by a large number of crystalline elements, such as aluminum, nickel, copper, silver, calcium, neon, argon and krypton. In addition, we find that the elements in column IV of the periodic table, silicon and germanium, crystallize in the same form as diamond carbon. In this structure, half of the atoms are in the same environment as those in a fcc structure. The other atoms are also in a fcc lattice that is displaced with respect to the former lattice by a quarter cell unit in the direction [1,1,1]. The two types of atoms differ simply in their orientation with respect to their bonds with their nearest neighbors.

Copper and silicon are both standard bearers in their respective classes of metals and semiconductors. Given their industrial and academic importance, it is easy to see why the study of their crystalline structures in order to understand their electronic properties is so crucial.

6.2. Direct and reciprocal lattices of the fcc structure

6.2.1. *Direct lattice*

The fcc structure (side *d*) is shown in Figure 6.1. Two forms of unit cell can be constructed:

i) A rhombohedric cell based on the vectors $\vec{a},\vec{b},\vec{c}$ joining the origin *O* to the center of the faces of the cube.

The component of the $\vec{a},\vec{b},\vec{c}$ vectors on the *x, y, z* axes are:

	x	y	z
$\vec{a}$	$d/2$	$d/2$	0
$\vec{b}$	$d/2$	0	$d/2$
$\vec{c}$	0	$d/2$	$d/2$

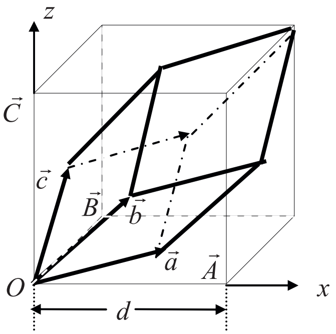


Figure 6.1. *Fcc structure*

The rhombohedron is the smallest cell possible; it is therefore called the primitive cell and has, on average, one node per cell (a node for each of the eight points of the rhombohedric cell, with each being shared with eight other rhombohedric cells).

ii) A cubic cell with the base vector being vectors $\vec{A},\vec{B},\vec{C}$ (from the *d* components respectively along the *x, y, z* axes) with the nodes at the summit of the cell and at the center of each *d* face side. This fcc cell contains, on average, four nodes per cell. There are eight nodes at the summit of each face that are each shared with eight other cells and that contribute, on average, one node per cell. However, there are also six nodes at the center of six faces that are each shared between six faces and therefore bring an extra 6/2 = 3 nodes per cell.

With a total of four nodes per cell, the centered cubic cell is four times larger than the rhombohedric primitive cell.

6.2.2. *Reciprocal lattice*

6.2.2.1. *Base vectors deduced directly from the direct lattice*

Using the primitive cell in the direct lattice as a starting point, the base vectors $\vec{A}, \vec{B}, \vec{C}$ of the reciprocal lattice are such that, for example with $\vec{a}^* = 2\pi \frac{\vec{b} \times \vec{c}}{(\vec{a}, \vec{b}, \vec{c})}$, we have:

$$\vec{b} \times \vec{c} \left\{ \begin{array}{l} -\dfrac{d^2}{4} \\ -\dfrac{d^2}{4} \\ \dfrac{d^2}{4} \end{array} \right. \text{ and } (\vec{a}, \vec{b}, \vec{c}) = \vec{a} \cdot \left(\vec{b} \times \vec{c} \right) = -\frac{d^3}{8},$$

so that:

$$\vec{a}^* \left\{ \begin{array}{l} 2\pi \left(-\dfrac{d^2}{4} \right) \left(-\dfrac{4}{d^3} \right) = \dfrac{2\pi}{d} \\ \dfrac{2\pi}{d} \\ -\dfrac{2\pi}{d} \end{array} \right.$$

Finally, by a circular permutation, we find the components of the three base vectors of the reciprocal lattice:

	x	y	z
$\vec{a}^*$	$\frac{2\pi}{d}$	$\frac{2\pi}{d}$	$-\frac{2\pi}{d}$
$\vec{b}^*$	$\frac{2\pi}{d}$	$-\frac{2\pi}{d}$	$\frac{2\pi}{d}$
$\vec{c}^*$	$-\frac{2\pi}{d}$	$\frac{2\pi}{d}$	$\frac{2\pi}{d}$

and the modules of these vectors are such that (see Figure 6.2):

$$\left|\vec{a}\,*\right| = \sqrt{\left(\frac{2\pi}{d}\right)^2 + \left(\frac{2\pi}{d}\right)^2 + \left(\frac{2\pi}{d}\right)^2} = \frac{2\pi}{d}\sqrt{3} = \left|\vec{b}\,*\right| = \left|\vec{c}\,*\right|$$

The vectors of the reciprocal lattice are in the form $\vec{G}_{h,k,l} = h\vec{a}\,* + k\vec{b}\,* + l\vec{c}\,*$, and the nodes closest to the origin are given by the vectors $\begin{cases}\vec{G}_{1,0,0} = \vec{a}\,* \\ \vec{G}_{0,1,0} = \vec{b}\,* \\ \vec{G}_{0,0,1} = \vec{c}\,*\end{cases}$. The Brillouin zones, which are in themselves difficult to geometricly construct, can be traced from the mediating planes of the vectors denoted $\vec{a}*$, $\vec{b}*$, $\vec{c}\,*$. An additional simplification is made by changing the base.

6.2.2.2. *Changing base in a reciprocal lattice*

A new base can be more easily constructed in a reciprocal lattice by using the new base vectors given by:

$$\begin{cases}\overrightarrow{A}\,* = \vec{a}\,* + \vec{b}\,* \\ \overrightarrow{B}\,* = \vec{a}\,* + \vec{c}\,* \\ \overrightarrow{C}\,* = \vec{b}\,* + \vec{c}\,*\end{cases}$$

In effect, the vectors $\overrightarrow{A}\,*$, $\overrightarrow{B}\,*$, $\overrightarrow{C}\,*$ make up a base as they correspond to the maximum number of linear vectors (equal to three). To confirm this property, we can create a linear combination given by $\mu_1\overrightarrow{A}\,* + \mu_2\overrightarrow{B}\,* + \mu_3\overrightarrow{C}\,*$, and show that if $\mu_1\overrightarrow{A}\,* + \mu_2\overrightarrow{B}\,* + \mu_3\overrightarrow{C}\,* = 0$, then $\mu_1 = \mu_2 = \mu_3 = 0$. Given the definition of $\overrightarrow{A}\,*$, $\overrightarrow{B}\,*$, $\overrightarrow{C}\,*$, the linear combination gives:

$$\mu_1\left(\vec{a}\,* + \vec{b}\,*\right) + \mu_2\left(\vec{a}\,* + \vec{c}\,*\right) + \mu_3\left(\vec{b}\,* + \vec{c}\,*\right) = 0$$

so that:

$$\vec{a}\,*\left(\mu_1 + \mu_2\right) + \vec{b}\,*\left(\mu_1 + \mu_3\right) + \vec{c}\,*\left(\mu_2 + \mu_3\right) = 0.$$

As $\vec{a}*,\ \vec{b}*,\ \vec{c}*$ are the base vectors, this equation implies that $(\mu_1+\mu_2)=(\mu_1+\mu_3)=(\mu_2+\mu_3)=0$, so that $\left.\begin{matrix}\mu_1=-\mu_2=-\mu_3\\ \text{and }\ \mu_2=-\mu_3\end{matrix}\right\}\Rightarrow \mu_1=\mu_2=\mu_3=0$, which in turn shows that $\overrightarrow{A*},\ \overrightarrow{B*},\ \overrightarrow{C*}$ are linearly independent.

Given the definition of the vectors (see Figure 6.2), the components of $\overrightarrow{A*}$, $\overrightarrow{B*}$, $\overrightarrow{C*}$ are thus:

	x	y	z
$\overrightarrow{A*}$	$\frac{4\pi}{d}$	0	0
$\overrightarrow{B*}$	0	$\frac{4\pi}{d}$	0
$\overrightarrow{C*}$	0	0	$\frac{4\pi}{d}$

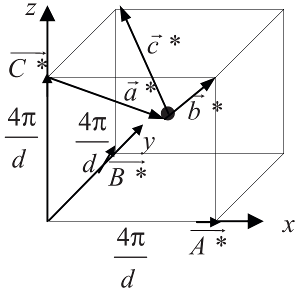


Figure 6.2. *Centered cubic (cc) of the reciprocal lattice of the fcc structure*

The new cell obtained for the reciprocal lattice is thus a cubic centered cell (ccc) with sides equal to $\frac{4\pi}{d}$, a node at each summit and a node at the center of the cube (see also the components of the primitive vectors of the reciprocal lattice $\vec{a}*,\ \vec{b}*,\ \vec{c}*$).

This is a classic result that can be stated simply enough, the reciprocal lattice of a fcc lattice is a cc lattice. Reciprocally, the reciprocal lattice of a cc lattice is a fcc.

6.3. Brillouin zone for the fcc structure

6.3.1. *Geometric form*

In the reciprocal lattice, the closest nodes to the origin are localized at:

– six coordinate points $\left(\pm\frac{4\pi}{d},0,0\right)$, $\left(0,\pm\frac{4\pi}{d},0\right)$, $\left(0,0,\pm\frac{4\pi}{d}\right)$, situated in the six directions normal to the planes [1,0,0], $[\bar{1},0,0]$, [0,1,0], $[0,\bar{1},0]$, [0,0,1], $[0,0,\bar{1}]$. The vector of the reciprocal lattice $\vec{G}_{1,0,0}$ normal to the planes [1,0,0] is such that $\vec{G}_{1,0,0}=\overrightarrow{A*}$, with $\left|\vec{G}_{1,0,0}\right|=\frac{4\pi}{d}$;

– eight cube centers from the combinations of $\left(\pm\frac{2\pi}{d},\pm\frac{2\pi}{d},\pm\frac{2\pi}{d}\right)$, and situated in the eight directions [1,1,1] (normal to planes [1,1,1], $[\bar{1},1,1]$, $[1,\bar{1},1]$, $[1,1,\bar{1}]$, $[\bar{1},\bar{1},1]$, $[\bar{1},1,\bar{1}]$, $[1,\bar{1},\bar{1}]$, $[\bar{1},\bar{1},\bar{1}]$). The vector of the reciprocal lattice normal to planes [1,1,1] and with an extremity at the center of the cube (in the first octant) is the vector:

$$\frac{\vec{G}_{1,1,1}}{2}=\frac{\overrightarrow{A*}}{2}+\frac{\overrightarrow{B*}}{2}+\frac{\overrightarrow{C*}}{2}.$$

Its module is such that:

$$\left|\frac{\vec{G}_{1,1,1}}{2}\right|=\sqrt{\left(\frac{2\pi}{d}\right)^2+\left(\frac{2\pi}{d}\right)^2+\left(\frac{2\pi}{d}\right)^2}=\frac{2\pi}{d}\sqrt{3}$$

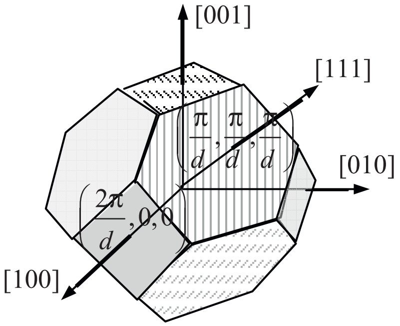


Figure 6.3. *First Brillouin zone of the fcc structure*

The first Brillouin zone described by the mediating planes of segments joining the origin node to its closest neighbors is shown in Figure 6.3. In the [1,0,0] directions the faces are squared and in the [1,1,1] directions they are regular hexagons. The coordinates of the intersection of the mediating planes with vectors $\vec{G}_{1,0,0}$ and $\vec{G}_{1,1,1}/2$ are indicated in the figure.

6.3.2. *Calculation of the volume of the Brillouin zone*

Figure 6.4 schematically illustrates the Brillouin zone (or more exactly an eighth of the Brillouin zone) in the first octant.

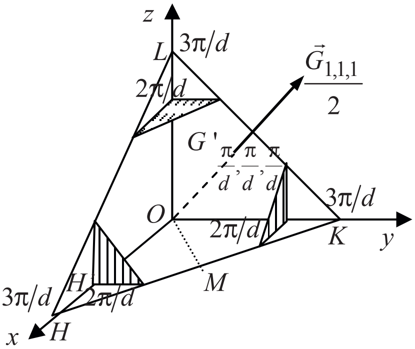


Figure 6.4. *Brillouin zone in the first octant*

In the [1,1,1] direction, the node is found at the extremity of vector $\frac{\vec{G}_{1,1,1}}{2}$, and the intersection with the mediating plane (Brillouin zone) of this segment is thus at the point denoted *G'* that has the co-ordinates $\left(\frac{\pi}{d},\frac{\pi}{d},\frac{\pi}{d}\right)$. The equation for this mediating plane can be found by considering a vector $\overrightarrow{V}$ (with components *x, y, z*) that has an extremity in the plane that is such that $\overrightarrow{V}\cdot\overrightarrow{OG'}=|\overrightarrow{OG'}|^2$, so that $x\frac{\pi}{d}+y\frac{\pi}{d}+z\frac{\pi}{d}=3\frac{\pi^2}{d^2}$, and hence $x+y+z=3\frac{\pi}{d}$.

The intersection of this plane with axis *x* is such that when $y=z=0$, we have $x=\frac{3\pi}{d}$ (the point *H* in Figure 6.4).

Similarly, the intersection between axes y and z occurs, respectively, at $y = \frac{3\pi}{d}$ and $z = \frac{3\pi}{d}$ (indicated as points L and K, respectively, in Figure 6.4). The points situated at $(2\pi/d, 0, 0)$, $(0, 2\pi/d, 0)$, $(0, 0, 2\pi/d)$ correspond for their part to the intersectional points at the x, y and z axes on the Brillouin surfaces, which were obtained as mediating planes in the [100], [010] and [001] directions.

The volume V' of the first Brillouin zone of the first Brillouin zone in the first octant thus takes on the form $V' = v - 3v_1$, where v is the volume enclosed by three axes denoted (Ox, Oy, Oz) and the plane defined by the intersectional points HKL, and v_1 represents each of the small triads at the point H, K and L of the bases clipped off in Figure 6.4. It is the removal of these pyramids that result in the truncated sides of the Brillouin zone. We thus find

– $v = \frac{1}{3} S \cdot h$, where $h = \frac{3\pi}{d}$ and $S = \frac{1}{2}\left(\frac{3\pi}{d}\right)^2$ so that $v = \frac{9\pi^3}{2d^3}$;

– $v_1 = \frac{1}{3} S_1 \cdot h_1$, where $h_1 = \frac{\pi}{d}$ and $S_1 = \frac{1}{2}\left(\frac{\pi}{d}\right)^2$ so that $v_1 = \frac{\pi^3}{6d^3}$, or rather, $3v_1 = \frac{\pi^3}{2d^3}$.

We thus find that $V' = v - 3v_1 = \frac{4\pi^3}{d^3}$, and that the total volume of the sliced octahedron, and therefore also of the first Brillouin zone is given by:

$$V = 8V' = \frac{32\pi^3}{d^3}$$

6.3.3. *Filling the Brillouin zone for a fcc structure*

If V_0 is the volume of the starting crystal (i.e. in direct space), then the number of (fcc) cells with a volume given by d^3 and containing an average of four nodes is given by $\frac{V_0}{d^3}$. In direct space we therefore have $N = 4\frac{V_0}{d^3}$ nodes.

The volume of a cell in reciprocal space is $\frac{8\pi^3}{V_0}$. In the volume V of the Brillouin zone of the fcc structure, we can therefore place:

$$\frac{V}{\frac{8\pi^3}{V_0}} = \frac{32\pi^3}{d^3}\frac{V_0}{8\pi^3} = \frac{4V_0}{d^3} = N \text{ cells}$$

that is, the same number as there are nodes in the direct space.

Being able to place two electrons per cell into the reciprocal space means that the fcc Brillouin zone structure will be filled with 2*N* electrons. However, this will only happen if the set of atoms (attached to each node) each liberate an even number of electrons.

6.4. Copper and alloy formation

6.4.1. *Electronic properties of copper*

The atomic number (*Z*) of copper is 29. Its electronic configuration is [Ar] $3d^{10}\,4s^1$. It is thus a mono-valent atom and therefore its fcc lattice carries *N* nodes on which are located *N* atoms that altogether liberate *N* electrons. The first Brillouin zone is thus half-filled meaning that copper is a metal.

We can also note that silver ($Z = 47$) and gold give rise to the same behavior. Their electronic configurations are, respectively, [Kr] $4d^{10}\,5s^1$ and [Xe], $4f^{14}\,5d^{10}\,6s^1$ ($Z = 79$).

The formation of alloys of copper with other more electronically rich elements is of particular interest given that copper has a Brillouin zone that is only half-full.

6.4.2. *Filling the Brillouin zone and solubility rules*

6.4.2.1. *Filling the Brillouin zone and the consequences*

At absolute zero, electrons occupy the minimum energy cells (starting from the origin of the reciprocal lattice) and are distributed inside the interior of the Fermi surface.

For the mono-valent elements (notably Cu, Ag, Au) that are crystallized in fcc structures, the number of cells occupied is considerably smaller than the number of cells that can be placed in the first Brillouin zone. The external surface of the distribution of full cells is very close to a sphere, and the scheme that represents the free electrons is very close to reality.

In contrast, if the element is richer in electrons, for example if it is divalent, then the first zone can be filled. However, as discussed in Chapter 5 for an equi-energy sphere that is tangential to the zone limit, there are two situations that can arise:

i) if there is a gap between the first and second zones, then the first zone fills up and the material exhibits insulator/semiconductor properties; or

ii) if the gap is small, then the second zone can start filling up before the first is completely filled, and this results in a semi-metallic behavior.

Thus, these two scenarios are separated by the point at which the electrons are just able to fill the sphere within the first Brillouin zone. The position of the equi-energy sphere, at a tangent to the zone limit, is important when evaluating the formation of alloys.

6.4.2.2. *Conditions for alloy formation: the Hume-Rothery rules*

6.4.2.2.1. Conditions for alloy formation

In order for alloys to be formed, the diameters of the atoms should be very close to one another, or otherwise a disparity in electrons can make it very difficult to form a solid solution. An excessive number of electrons can, however, make alloy formation quite easy. For example, mono-valent copper is only slightly soluble (1%) in bivalent zinc ($Z = 30$) which has the electronic structure [Ar] $3d^{10}\ 4s^2$. But zinc is highly soluble in copper (up to 1/3 zinc and 2/3 copper). Similarly, copper is insoluble in silicon (valency of four), whereas silicon can be placed into copper (up to 6%).

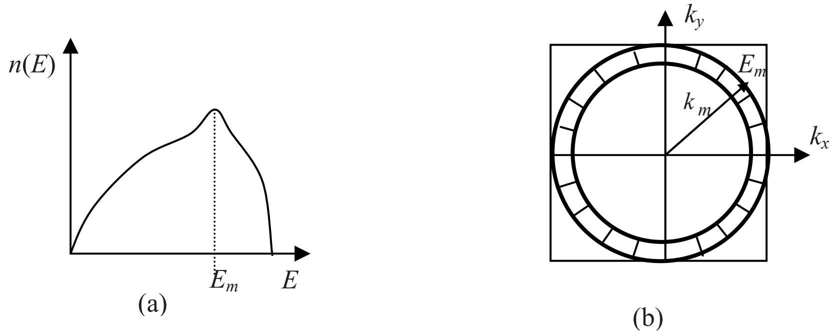


Figure 6.5. *(a) Variation in the maximum energy of n(E) and denoted as E_m; (b) scheme showing how the greatest number of cells filled by electrons with a given energy E is obtained with an equi-energy sphere that is tangential to the Brillouin zone*

In addition, we can note that the maximum of the curve $n(E)$ (which describes $n(E)= Z(E)f(E)$ for the electronic density) gives the maximum value of the attained electronic energy (E_m) (see also Figure 4.7).

Given the remarks above, we can easily see that this energy value relates to the position of the equi-energy sphere tangential to the zone limit. In effect, it is at a position where the sphere covers the greatest surface that can surround the maximum number of cells that accommodate electrons of the same energy E (which is equal to the maximum value of $n(E)$ for $E = E_m$). At a smaller radius, the surface is smaller, but at higher values the surface is broken by the limits of the zone; see Figure 6.5b, which sketched a simplified example of a 2D square lattice that results in a square Brillouin zone; and Chapter 5, where the radius of the circle tangential to the zone limit is equal to:

$$k_m = \frac{\sqrt{2mE_m}}{\hbar}$$

This concentration of electrons at the maximum of $n(E)$ is very important in metallurgy. If we increase the electronic concentration above this limit, then the energy of the structure increases considerably (as the gap at the zone limit must be crossed) and becomes unstable.

6.4.2.2.2. The Hume-Rothery solid solubility rules

These rules are deduced from the preceding results and the resulting general rule: the electronic concentration observed for a stable alloy is given by the Fermi sphere tangential to the Brillouin zone limit.

NOTE.– As above, the value given by $E = E_m$ the structure becomes unstable, the value of $E = E_m$ becomes the maximum energy attained and is therefore the Fermi energy (E_F) of the alloy.

RESULT.– For a fcc lattice the limiting (maximum) electronic concentration is 1.36 electrons per atom (see section 6.4.3). For a cc structure, the limiting concentration is around 1.48 electrons per atom (see problems, section 6.6).

6.4.2.2.3. Simple examples

For a square lattice (2D) of length and unit spacing denoted L and a, respectively, the surface of the direct lattice is $S = L^2$ and the cell of the primitive cell is given by a^2. The number of cells is thus given by $N = L^2/a^2$ which is also the number of nodes (there is on average one node per cell). In the reciprocal lattice, the surface of the primitive cell is given by $(2\pi/L)^2$.

The Brillouin zone is a square of side $(2\pi/a)$ and the (maximum) radius of the equi-energy circle is equal to (π/a). Within this circle that has a surface of $\pi(\pi/a)^2$ we can place:

$$\frac{\pi^3/a^2}{(2\pi/L)^2} = \frac{\pi}{4}\frac{L^2}{a^2} = \frac{\pi}{4}N \text{ cells,}$$

or rather $\frac{\pi}{2}N = 1.57 \cdot N$ electrons. If the atomic base is equal to 1, that is, one atom at each node, then there are N atoms in all with a concentration of electrons equal to 1.57 electrons per atom.

For a cubic lattice of side L and unit spacing a, the volume of the direct lattice is $V = L^3$ and that of a primitive cell is equal to a^3. The number of cells is thus $N = L^3/a^3$ which is also the number of nodes (an average of one node per cell). In the reciprocal lattice, the volume of the primitive cell is $(2\pi/L)^3$, the Brillouin zone is a cube with sides equal to $(2\pi/a)$, and the (maximum) radius of the equi-energy sphere is (π/a). Within the volume of the sphere $\frac{4}{3}\pi\left(\frac{\pi}{a}\right)^3$ we can place $\frac{4}{3}\pi\left(\frac{\pi}{a}\right)^3 \Big/ \frac{8\pi^3}{L^3} = \frac{\pi}{6}\frac{L^3}{a^3} = \frac{\pi}{6}N$ cells, so that there are $\frac{\pi}{3}N = 1.05 \cdot N$ electrons. If we again have an atomic base equal to one (that is one at each node), we find that for the total N atoms the electronic concentration is equal to 1.05 electrons per atom.

6.4.3. *Copper alloys*

This section looks at copper alloy formed with electronically richer atoms (for example bivalent atoms) and uses the same notation as in section 6.2. A stable structure is obtained when the equi-energy sphere is at a tangent to the limit of the zone. This happens when this sphere centered at the origin reaches the point G' shown in Figure 6.4, so that $OG' = \frac{\pi}{d}\sqrt{3}$. The volume of the sphere is now given by $V_s = \frac{4}{3}\pi\left(\frac{\pi}{a}\sqrt{3}\right)^3 = \frac{4\sqrt{3}\pi^4}{a^3}$ and a number $N' = \frac{V_s}{\frac{8\pi^3}{V_0}} = \frac{\pi\sqrt{3}V_0}{2d^3}$ of cells of volume $\left(\frac{8\pi^3}{V_0}\right)$ can be placed within the sphere. It should be noted that $V_0 = \frac{d^3}{4}N$ (see also section 6.3.3 where N is the number of nodes that is in this case the number of atoms in the direct space as only a single atom can occupy each node, be it copper or the added metal). The number of electrons that fill this volume is now given by $2N' = \frac{\pi\sqrt{3}}{4}N = 1.36 \cdot N$. In other words, for the sphere to be thought of as being just-filled, the electronic concentration should be at 1.3 electrons per atom. This is the sort of level of electron density that can be attained in copper alloys when zinc is added and an alpha (α) phase alloy is acquired. A higher level can be attained when the alloy takes up a cubic centered beta (β) phase with around 1.48 electrons per atom (see the problem at the end of the section 6.6, question 5).

6.5. Silicon

6.5.1. *The silicon crystal*

Silicon is classed as the 14th element in the periodic table and it possesses four peripheral electrons in the *M* layer. In a silicon crystal, each atom is engaged in four bonds with four neighboring atoms. The silicon crystal is thus covalent and the four bonds take up tetrahedral positions (with an angle of 109° 28’) to give a structure not unlike that of diamond.

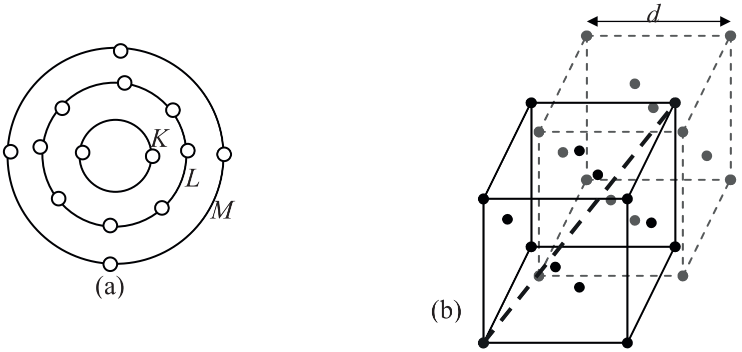


Figure 6.6. *Silicon’s: (a) electronic structure; and (b) crystal cell*

Silicon atoms are placed at the nodes of two cubic face centered lattices that are shifted with respect to one another by a quarter cube diagonal. This results in the cell shown in Figure 6.6b, which can also be used to represent diamond carbon (hence its common name, the “diamond lattice”) or germanium crystals. In the case of silicon, the primitive cube has a lattice constant (*d*) equal to 5.43 Å, whereas for carbon, $d = 3.56$ Å and for germanium $d = 5.62$ Å. Each primitive fcc cell with a volume of

d^3 contains on average eight atoms: there are four belonging to one fcc lattice and four belonging to the other. We can assume that on average that there are four atomic bases per fcc cell. With two atoms per base, there are in all eight atoms per fcc cell.

6.5.2. *Conduction in silicon*

Given its structure, and that it crystallizes in an fcc system, with two silicon atoms per node (base = two silicon atoms so as to account for the two imbricating lattices), the N nodes are for $2N$ silicon atoms. Given that their valency is equal to four, they liberate $8N$ electrons which can fill several successive Brillouin zones (each with $2N$ electrons). The bands do not overlap and the width of the forbidden band has been determined as being $E_G = 1.12$ eV, indicating that silicon is a semiconductor. The size of forbidden band decreases slightly as temperature increases due to the dilatation of the crystalline structure, so that at 100°C, we find that $E_G = 1.09$ eV. Of note is that the forbidden band width of germanium is 0.66 eV.

6.5.3. *The silicon band structure (see also section 8.4)*

Determining the band structure of a material means finding the correspondence between the wave vector and the energy at all points in different zones (in general reduced to the first zone) or the different bands.

For a 1D material, the energy of the semi-free electrons is given by equation [4.10] in section 4.3 that states $E_{(k)} = \mathrm{E}_{(\mathrm{k}_\mathrm{n})} + \frac{\hbar^2}{2\mathrm{m}^*}(k - k_n)^2$, where $k_0\,n$ is the value of k at the limit of the zone which has remained till now in the form $k_n = n\frac{\pi}{a}$. For a 3D material (where the transport and therefore the effective mass can differ with respect to the x, y and z directions), in the preceding expression m^* can take on various values, namely: m_x^*, m_y^* and m_z^* for the respective directions x, y and z. In addition, it is possible that the extreme values for the energy are not obtained at the band limits (in the direction k_x for $k_x = k_{nx}$ and similarly for k_y and k_z) but at other points in the k_x, k_y or k_z directions.

6.5.3.1. *Properties of the conduction band*

We thus find that silicon shows a conduction band in the [100] direction with a minimum at $k_m = 0.85\frac{2\pi}{a}$, where $\frac{2\pi}{a}$ is the limit of the Brillouin zone in these directions (see Figure 6.3 and 6.4). As there are six [100] type directions, there are six minimum equivalents, and in the neighborhood of k_m the energy is given by:

$$E = E_C + \frac{\hbar^2}{2m_\ell^*}\left(k_x - k_m\right)^2 + \frac{\hbar^2}{2m_t^*}\left(k_y^2 + k_z^2\right)$$

with

$$m_\ell^* = \frac{\hbar^2}{\dfrac{\partial^2 E}{\partial k_x^2}} \text{ and } m_t^* = \frac{\hbar^2}{\dfrac{\partial^2 E}{\partial k_y^2}} = \frac{\hbar^2}{\dfrac{\partial^2 E}{\partial k_z^2}}$$

In practical terms, the values of m_ℓ^* and m_t^* are obtained through cyclotron resonance characterizations (performed using a magnetic field on conduction electrons). With m_0 being the electron rest mass, we find that:

$$m_\ell^* \approx 0.90\, m_0 \text{ and } m_t^* \approx 0.19\, m_0$$

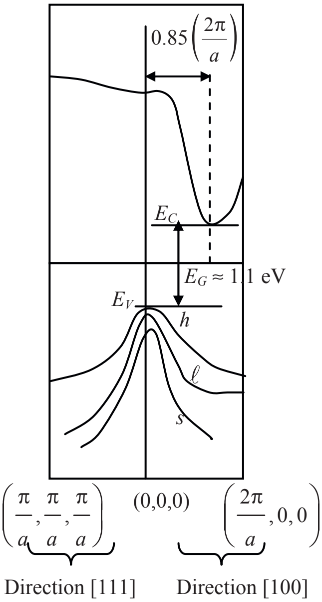


Figure 6.7. *Correspondence between wave vector and energy in the [100] and [111]*

6.5.3.2. *Properties of the valence band*

The valence band shows a maximum when $k = 0$. Outside of this value of the wave vector, the band is separated into three curves due to spin-orbital interactions giving rise to the h, ℓ and s bands. In the case of silicon, these arise from the $3p^2$ electronic configuration states that are in the outer $3s^2$ $3p^2$ layer. To the $3p^2$ states there are two types of electrons that are associated with two j values, namely:

– $j = \frac{3}{2}$ states obtained when $j = \ell + s$ that give rise to four states characterized by $2j+1$ values of m_j. This means that we have $m_j = -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}$. The four states denoted $p_{3/2}$ have greater energies than the following states (denoted $p_{1/2}$) and give rise to bands denoted h and ℓ (degenerated as the two bands are for four states). With the band having an acute maximum, it is populated with light holes (hence the name ℓ-band for light hole) as shown in Figure 6.7, while the other gives rise to heavy holes (hence the name h-band).

– $j = \frac{1}{2}$ states obtained when $j = \ell - \frac{1}{2}$ that give rise to the s band (see Figure 6.7) which has a parabolic character. This band is also degenerate as there are two states for each m_j, as $m_j = -\frac{1}{2}, \frac{1}{2}$.

6.5.3.3. *Consequences expressed in opto-electronic-properties*

When the extremes of the valence and conduction bands are obtained for the same value of k (see Figure 6.8b) we have a direct band structure that has a direct (vertical) transition between the lowest point of the conduction band and the highest of the valence band. The direct transition is thus one where there is only a variation in the energy (between E_C at the base of the conduction band and E_V at the summit of the valence band we have $\Delta E = E_C - E_V$), and without variation in k (E_C and E_V are levels with the same value in k terms, i.e. k_0).

However, in the case of silicon, we have a material which exhibits a so-called indirect gap (see Figure 6.8a) and the transition between the lowest point in the conduction band the summit of the valence band entails both a variation in energy given by $\Delta E = E_C - E_V$ and a change in k. The extrema of the conduction (E_C) and valence (E_V) bands are obtained at different values of k so that the transition involves a change in k given by $\Delta k = k_C - k_V$.

While a photon suffices to give a change in energy in a vertical transition, an indirect transition needs the simultaneous intervention of another particle which will give (absorption) or receive (emission) the variation in k given by Δk. These vibration levels (represented by a *quasi*-particle called a phonon; see Chapter 10) ensure this otherwise unlikely transition, unlikely as it necessitates the concomitant intervention of three particles: electron, photon and phonon (see the second volume of *Materials and Electronics* for more information).

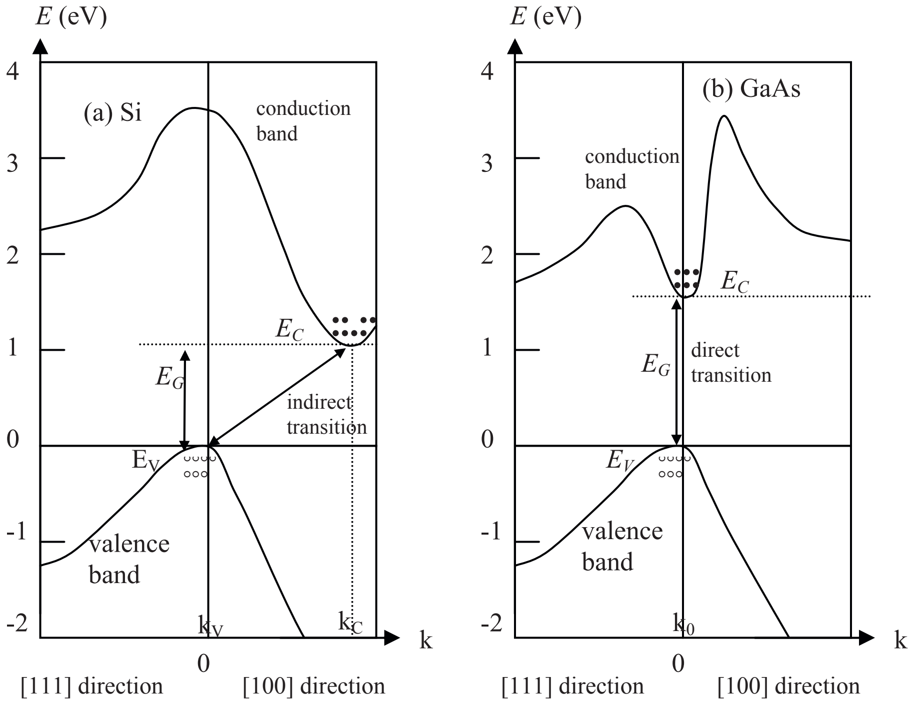


Figure 6.8. *Band structures: (a) with an indirect gap (as in Si); (b) with a direct gap (as in GaAs)*

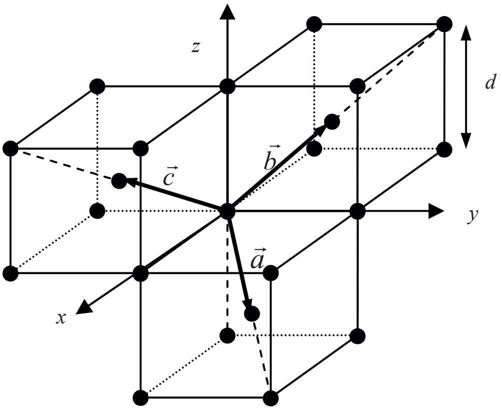
6.5.4. *Conclusion*

The structure of the silicon band shows how this material is particularly important in electronics as the size of its gap (1.1 eV) is well adapted to the semiconductive filling of its bands (see the second volume of materials and electronics for more information). However, the indirect structure of its gap means that it is not so useful for opto-electronic applications, a field in which it is bettered – by other semiconductors such as GaAs (and the materials in columns III–V of the periodic table) that have direct gaps.

6.6. Problems

6.6.1. *Problem 1: the cubic centered (cc) structure*

A cc lattice is shown in the figure below.



	x	y	z
$\vec{a}$	$d/2$	$d/2$	$-d/2$
$\vec{b}$	$-d/2$	$d/2$	$d/2$
$\vec{c}$	$d/2$	$-d/2$	$d/2$

With respect to axes x, y and z, the fundamental vectors of the direct lattice ($\vec{a},\vec{b},\vec{c}$) have the following components where d is the side of a conventional cube as shown in the figure:

1) Determine the fundamental vectors $\vec{A},\vec{B},\vec{C}$ of the reciprocal lattice. Give the structure of the reciprocal lattice of the cc direct lattice.

2) What are the 12 small vectors of the reciprocal lattice?

3) Deduce the form of the first Brillouin zone of the cc lattice.

4) Calculate the total volume of this first Brillouin zone.

5) Determine the number of electrons per atom that must be present in order to just fill this Brillouin zone (assuming that the atomic base is equal to 1, i.e. there is one atom at each node).

6) And alloy of two crystalline materials is prepared in a cubic centered system. What is the optimum number of electrons per atom to give a stable alloy?

Answers

1) We can use for example $A = 2\pi\frac{\vec{b}\times\vec{c}}{V}$, where $V = \vec{a}(\vec{b}\times\vec{c}) = \frac{d^3}{2}$ is the volume of a cubic cell. With $\left(\vec{b}\times\vec{c}\right)$ that has components ($d^2/2$, $d^2/2$, 0), we thus obtain:

$$\vec{A}\begin{bmatrix}\frac{2\pi}{d}\\ \frac{2\pi}{d}\\ 0\end{bmatrix}\quad \vec{B}\begin{bmatrix}0\\ \frac{2\pi}{d}\\ \frac{2\pi}{d}0\end{bmatrix}\quad \vec{C}\begin{bmatrix}\frac{2\pi}{d}\\ 0\\ \frac{2\pi}{d}\end{bmatrix}$$

These vectors exhibit the same components as the fundamental vectors of a face cc lattice for which each side of the cube is worth $\frac{4\pi}{d}$ (rather than just d as is the case in the direct lattice as shown in section 6.2.1). The up-shot is that the structure of the reciprocal lattice (of the cc direct lattice) is fcc.

2) The vectors of the reciprocal lattice are defined by:

$$\begin{aligned}\vec{G}_{h,k,l} &= h\vec{A} + k\vec{B} + l\vec{C}\\ &= \frac{2\pi}{d}\left[\left(h+l\right)\vec{e}_x + \left(h+k\right)\vec{e}_y + \left(k+l\right)\vec{e}_z\right].\end{aligned}$$

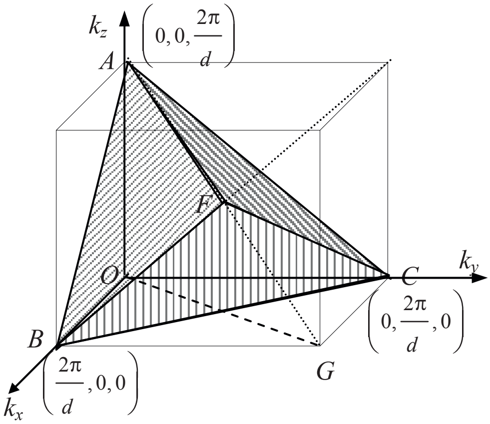
The smallest vectors ($\vec{G}$) are vectors given by the following expressions, in which the signs vary independently from one another:

$$\frac{2\pi}{d}\left(\pm\vec{e}_x \pm\vec{e}_y\right),\ \frac{2\pi}{d}\left(\pm\vec{e}_y \pm\vec{e}_z\right),\ \frac{2\pi}{d}\left(\pm\vec{e}_x \pm\vec{e}_z\right).$$

3) The first Brillouin zone is delimited by the mediating planes of the preceding 12 $\vec{G}$ vectors. The vectors that join the origin to the centers of the faces of these zones are one-half of the preceding vectors:

$$\frac{\pi}{d}\left(\pm\vec{e}_x \pm\vec{e}_y\right),\ \frac{\pi}{d}\left(\pm\vec{e}_y \pm\vec{e}_z\right),\ \frac{\pi}{d}\left(\pm\vec{e}_x \pm\vec{e}_z\right)$$

We thus find ourselves with a regular dodecahedron. In an octant there are three faces given by *AFC*, *BFC* and *AFB* (see the figure). In all there are 3 × 8 = 24 faces, each sharing an adjacent octant so that there are in effect 12 faces throughout all space.



4) The volume in the first octant of the first Brillouin zone can be obtained as a difference between the volume of the pyramid *ABGCO* and the pyramid *FBGC*.

Thus, with:

$$V_{\text{ABGCO}} = \frac{1}{3}\left(\frac{2\pi}{d}\right)^2 \frac{2\pi}{d} = \frac{8\pi^3}{3d^3} \text{ and } v_{\text{FBGC}} = \frac{1}{3}\left(\frac{1}{2}\left[\frac{2\pi}{d}\right]^2\right)\frac{\pi}{d} = \frac{2\pi^3}{3d^3}$$

we find the Brillouin zone volume in the first octant to be:

$$V_{\text{ZB1Oc}} = \frac{8\pi^3}{3d^3} - \frac{2\pi^3}{3d^3} = \frac{2\pi^3}{d^3}$$

from which the total volume of the Brillouin zone over all space is given by:

$$V = 8V_{\text{ZB1Oc}} = \frac{16\pi^3}{d^3}$$

5) In a direct lattice with an atomic base equal to one atom per node, a cell with volume d^3 contains on average two atoms per cell. The eight atoms situated at the eight corners of the cube are each shared between eight cells altogether bring one

atom per cell, and the atom at the center of the cell that is not shared with any other cell also brings one atom per cell, thus giving a total of two atoms per cell. If the volume of the material is equal to V, there are $\frac{V}{d^3}$ cells in direct space, so that there are $\frac{2V}{d^3} = N$ atoms in the direct lattice.

In the reciprocal space, the volume of a cell is equal to $\frac{8\pi^3}{V}$. Therefore, in the Brillouin zone we can place $\left[\frac{16\pi^3}{d^3} \Big/ \frac{8\pi^3}{V}\right] = \frac{2V}{d^3} =$ N cells. There are therefore in the first Brillouin zone as many cells as atoms in the direct lattice, so all cells will be filled (with two electrons) if each atoms liberates two electrons. The Brillouin zone is thus totally filled if the atoms are divalent (a result identical to that found for a simple cubic or fcc lattice).

6) A stabilization of an alloy based on monovalent atoms (which by themselves can only fill half of the Brillouin zone) using divalent atoms is obtained when the equi-energy sphere tangential to the zone limit is filled with electrons. This means that the radius of this sphere is equal to:

$$\frac{|OG|}{2} = \frac{1}{2}\frac{2\pi}{d}\sqrt{2} = \frac{\pi}{d}\sqrt{2} = R$$

and that its volume is given by:

$$V_{\mathrm{F}} = \frac{4}{3}\pi R^3 = \frac{4}{3}\frac{\pi^4}{d^3}2\sqrt{2}$$

The sphere thus contains $\left[\frac{4}{3}\frac{\pi^4}{d^3}2\sqrt{2} \Big/ (8\pi^3/V)\right] = \frac{\pi\sqrt{2}}{3d^3}V = N'$ cells whch is equal to $2N'$ electrons when filled. Under these conditions and with a population of N atoms, we have $2N'/N$ electrons per atom so that there are $\left[2\frac{\pi\sqrt{2}}{3d^3}V \Big/ \frac{2V}{d^3}\right] = \frac{\pi\sqrt{2}}{3} = 1.48$ electrons per atom.

6.6.2. *Problem 2: state density in the silicon conduction band*

To determine the form of the state density function for the silicon conduction band, we should note that in the neighborhood of $\vec{k}_m$ (wave vector for a minimum in the conduction band where the minimum energy $E_{\min} = E_C$), we can write that the energy of a semi-free electron (of non-isotropic effective mass) is given by:

$$E = E_C + \frac{\hbar^2}{2m_\ell^*}(k_x - k_m)^2 + \frac{\hbar^2}{2m_t^*}\left(k_y^2 + k_z^2\right)$$

where:

$$m_\ell^* = \frac{\hbar^2}{\dfrac{\partial^2 E}{\partial k_x^2}} \text{ and } m_t^* = \frac{\hbar^2}{\dfrac{\partial^2 E}{\partial k_y^2}} = \frac{\hbar^2}{\dfrac{\partial^2 E}{\partial k_z^2}}$$

1) When introducing the vectors:

$$\vec{k'}\begin{cases} k'_x = k_x\sqrt{\dfrac{m^*}{m_\ell^*}} \\ k'_y = k_y\sqrt{\dfrac{m^*}{m_t^*}} \\ k'_z = k_z\sqrt{\dfrac{m^*}{m_t^*}} \end{cases} \quad \text{and} \quad \overrightarrow{k'_m}\begin{cases} k'_{mx} = k'_m = k_m\sqrt{\dfrac{m^*}{m_\ell^*}} \\ k'_{my} = 0 \\ k'_{mz} = 0 \end{cases}$$

what is the form of $E = f(E_C, \vec{k'}, \overrightarrow{k'_m}, m^*)$?

2) The reasoning used is with respect to *k'* space, but what is the dimension of a primitive cell is this space? What is the equi-energy surface in this *k'* space?

3) Determine the new expression for the state density function. What happens to this expression on taking into account the fact that the conduction band in reality displays six minima corresponding to six [100]-type directions that are equivalent and shown in Figure 6.3 (limited to the first octant and the equivalent directions [100], [010] and [001])?

4) Show how we are driven to defining a new effective mass (called the electron state density effective mass). How can it be expressed as a function of m_ℓ^* and m_t^* ?

Given that $m_\ell^* = 0.90\, m_0$ and $m_t^* = 0.192\, m_0$ (with $m_0 = 0.9 \times 10^{-30}$ kg), use the new expression for the effective mass as a function of m_0.

Answers

1) We have $E = E_C + \frac{\hbar^2}{2m_\ell^*}\left(k_x - k_m\right)^2 + \frac{\hbar^2}{2m_t^*}\left(k_y^2 + k_z^2\right)$, so that

– either $k_x = k_x' \sqrt{\frac{m_\ell^*}{m^*}}$ and $k_m = k_m' \sqrt{\frac{m_\ell^*}{m^*}}$, from which $(k_x - k_m)^2 = \frac{m_\ell^*}{m^*}(k_x' - k_m')^2$.

– or $k_y^2 = \frac{m_t}{m^*} k_y'^2$ and $k_z^2 = \frac{m_t}{m^*} k_z'^2$, in which case

$$E = E_C + \frac{\hbar^2}{2}\left[\frac{1}{m_\ell^*}\frac{m_\ell^*}{m^*}\left(k_x' - k_m'\right)^2 + \frac{1}{m_t^*}\frac{m_t^*}{m^*} k_y'^2 + \frac{1}{m_t^*}\frac{m_t^*}{m^*} k_z'^2\right]$$

$$= E_C + \frac{\hbar^2}{2m^*}\left[\left(k_x' - k_m'\right)^2 + k_y'^2 + k_z'^2\right]$$

By introducing $\vec{k}' \begin{cases} k_x' \\ k_y' \\ k_z' \end{cases}$ and $\overrightarrow{k_m'} \begin{cases} k_m' \\ 0 \\ 0 \end{cases}$, we finally reach:

$$E = E_C + \frac{\hbar^2}{2m^*}\left(\overrightarrow{k'} - \overrightarrow{k_m'}\right)^2$$

2) In reciprocal space, the dimension of a cell is $\Delta k_x \cdot \Delta k_y \cdot \Delta k_z = \Delta k^3 = \frac{8\pi^3}{V}$ (where V is the volume of the direct lattice). From:

$$\begin{cases} k_x' = k_x \sqrt{\dfrac{m^*}{m_\ell^*}} \\ k_y' = k_y \sqrt{\dfrac{m^*}{m_t^*}}, \\ k_z' = k_z \sqrt{\dfrac{m^*}{m_t^*}} \end{cases}$$

we deduce that:

$$\begin{cases} \Delta k_x' = \Delta k_x \sqrt{\dfrac{m^*}{m_\ell^*}} \\ \Delta k_y' = \Delta k_y \sqrt{\dfrac{m^*}{m_t^*}}, \\ \Delta k_z' = \Delta k_z \sqrt{\dfrac{m^*}{m_t^*}} \end{cases}$$

from which the dimension of a primitive cell in $\overrightarrow{k}'$ space is:

$$\Delta k_x' \cdot \Delta k_x' \cdot \Delta k_x' = \Delta k_x \cdot \Delta k_x \cdot \Delta k_x \sqrt{\frac{m^*}{m_\ell^*}\frac{m^*}{m_t^*}} = \frac{8\pi^3}{V}\frac{(m^*)^{3/2}}{\left(m_\ell^*\right)^{1/2} m_t^*}.$$

The relation obtained from question 1, i.e. $E = E_C + \frac{\hbar^2}{2m^*}\left(\overrightarrow{k'} - \overrightarrow{k_m'}\right)^2$, makes it possible to state that the equi-energy surface is a sphere centered at k_m' and with radius given by $\left|k' - k_m'\right| = \frac{\sqrt{2m^*|E - E_C|}}{\hbar}$.

3) In the k' space, the volume corresponding to the energies between E and (E + dE) is given by that between two spheres of radius $\left(\left|k' - k_m'\right|\right)$ and $\left(\left|k' - k_m'\right| + dk'\right)$, in other words a volume $dV_{k'} = 4\pi\left(\left|k' - k_m'\right|\right)^2 dk'$. From $\left|k' - k_m'\right|^2 = \frac{2m^*|E - E_C|}{\hbar^2}$, we deduce that $2\left|k' - k_m'\right| dk' = \frac{2m^* dE}{\hbar^2}$, and then that:

$$\left|k' - k_m'\right|^2 dk' = \frac{|m^*|\sqrt{2m^*|E - E_C|}}{\hbar^3} dE.$$

In turn from this we have:

$$dV_{k'} = 4\pi \frac{|m^*|\sqrt{2m^*|E - E_C|}}{\hbar^3} dE$$

The number of cells that can be placed in this volume is given by:

$$\frac{dV_{k'}}{\Delta k'_x \, \Delta k'_y \, \Delta k'_z} = 4\pi \frac{m^* \sqrt{2m^*|E - E_C|}}{\hbar^3} dE \cdot \frac{V \left(m_\ell^*\right)^{1/2} m_t^*}{8\pi^3 \left(m^*\right)^{3/2}}$$
$$= 4\pi \frac{\sqrt{2|E - E_C|}}{h^3} \left(m_\ell^*\right)^{1/2} m_t^* \; V \; dE$$

Finally, relative to the unit volume of the material ($V = 1$), we can place between the energy surfaces E and $E + dE$ a number denoted $Z(E)\,dE$ of electrons. This is given by (with two electrons per cell):

$$Z(E)dE = \frac{4\pi}{h^3} 2^{3/2} \left(m_\ell^*\right)^{1/2} m_t^* \sqrt{|E - E_C|} \; dE$$

Taking the six minima into account, the state density function can be written as:

$$Z(E) = 6 \frac{4\pi}{h^3} 2^{3/2} \left(m_\ell^*\right)^{1/2} m_t^* \sqrt{|E - E_C|}$$

4) If $m_c = \left(6[m_\ell^*]^{1/2} m_t^*\right)^{2/3}$, we have $Z(E) = \frac{4\pi}{h^3}\left(2m_c\right)^{3/2} |E - E_C|^{1/2}$, which is an equation analogous to that obtained classically for 3D space. With $m_\ell^* = 0.90\, m_0$ and $m_t^* = 0.192\, m_0$, we find $m_c = 1.05\, m_0$.

Chapter 7

Strong Bonds in One Dimension

This chapter starts with a concise description of the origin and construction of atomic and molecular orbitals found in a covalent solid (molecular films and polymers included). The results are then applied to energy levels in 1D covalent materials (notably molecular wires).

7.1. Atomic and molecular orbitals

7.1.1. *s- and p-type orbitals*

In the approximation for an atomic configuration (that gives the quantum numbers n, l, m…), we assume that each electron of an atom moves in a potential that has a spherical symmetry. The result is that:

– the potential of the nucleus varies with respect to $1/r$;

– this spherical potential gives a first approximation to the action of the other electrons.

The electronic state is thus represented by a wave function denoted $\psi_{n,l,m}$ that is dependent on three quantum numbers n, l and m, while the energy is only dependent on n and l (the degree of degeneration is equal to the number of values that m can take on). More details on this can be found in most basic courses on wave and atomic physics that use hydrogen as an example of a system with a spherical potential symmetry.

When $l = 0$, the atomic orbitals are denoted by the letter s and the wave functions only depend on n: $\Psi_{n,l,m} = R_{(r)}\,\Theta_{(\theta)}\,\Phi_{(\varphi)} = R_{n,l}\,\Theta_{l,m}\,\Phi_m = R_{n,0}\,\Theta_{0,0}\,\Phi_0$, where:

$$\Theta_{0,0} = \frac{\sqrt{2}}{2} \quad \text{and} \quad \Phi_0 = \frac{1}{\sqrt{2\pi}}$$

These wave functions do not depend on θ nor φ; only $R_{(r)}$ takes only different values as n, the principal quantum number, varies. The s-orbital thus has a spherical symmetry (as shown in Figure 7.1). The general expression for the radial functions is given by $R_{n,l}(r) = N_{nl}\left(\frac{2r}{na_1}\right)^l \exp\left(-\frac{r}{na_1}\right) L_{n+l}^{2l+1}\left(\frac{2r}{na_1}\right)$ where L represents the Laguerre polynomials and a_1 is the first Bohr radius (0.53 Å). The fundamental state is written using $R_{0,1} = Ce^{-r/a_1}$.

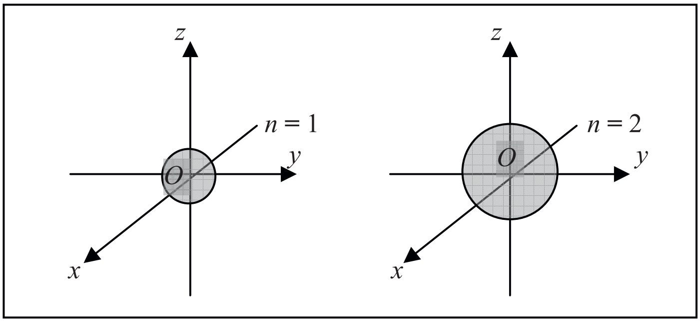


Figure 7.1. *Representation of s orbitals*

When $l = 1$, $m = -1, 0, 1$ and the orbitals are denoted by the letter p:

$m = 0$: $\Theta_{1,0} = (\sqrt{6}/2)\cos\theta$ and $\Phi_0 = 1/\sqrt{2\pi}$
$m = 1$: $\Theta_{1,1} = (\sqrt{3}/2)\sin\theta$ and $\Phi_1 = (1/\sqrt{\pi})\cos\varphi$
$m = -1$: $\Theta_{1,-1} = (\sqrt{3}/2)\sin\theta$ and $\Phi_{-1} = (1/\sqrt{\pi})\sin\varphi$

If $R_{(r)}$ represents the $R_{n,1}$ function, which does not change as m changes, the wave functions for the three preceding p states are of the form:

$$\Psi_0 = \frac{1}{2}\sqrt{\frac{3}{\pi}}R(r)\ \cos\theta \qquad [7.1]$$

$$\Psi_1 = \frac{1}{2}\sqrt{\frac{3}{\pi}}R(r)\ \sin\theta\ \cos\phi \qquad [7.2]$$

$$\Psi_{-1} = \frac{1}{2}\sqrt{\frac{3}{\pi}}R(r)\ \sin\theta\ \sin\phi \qquad [7.3]$$

with $R(r) = R(r_f)$ such that r_f has a value defined by the relation:

$$\int_0^{r_f} |R(r)|^2\ r^2\, dr = 95\%$$

(that indicates that the probability of the presence of an electron being inside the sphere of radius r_f is equal to 95%), functions [7.1]–[7.3] give a conventional representation of the orbitals.

NOTE.– The position of the M points are such that $\Psi = OM$, as shown in Figure 7.2.

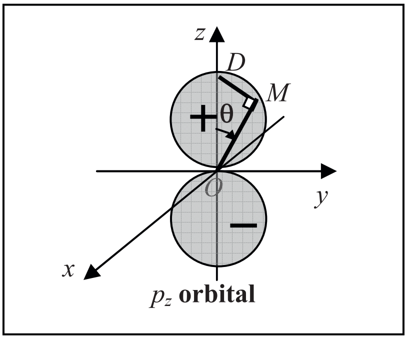


Figure 7.2. *The shape of the p_z orbital. The + and – signs give the sign of Ψ_0 which is that of cos θ*

We can use as a simple example the case where:

$$\Psi_0 = \frac{1}{2}\sqrt{\frac{3}{2}}R(r)\cos\theta = \left\|\overrightarrow{OM}\right\| = OM.$$

Here D is the point on the Oz axis where $\theta = 1$, and therefore we have:

$$OD = \frac{1}{2}\sqrt{\frac{3}{\pi}}R(r)$$

In other words $OM = OD \cos\theta$. If $\theta \in [0, \pi/2]$, we have $OM > 0$. The position of M points, as θ varies, is thus that of a sphere placed above the plane (xOy). If $\theta \in [\pi/2, \pi]$, $\cos\theta < 0$, and the M points are placed like a symmetric sphere below the plane (xOy).

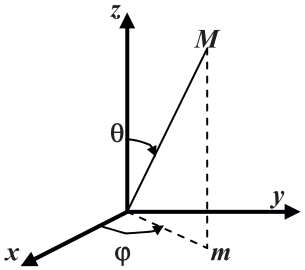


Figure 7.3. *Usually chosen spherical coordinates*

In addition, the signs of the Ψ_0, Ψ_1 $\Psi_{-1,}$ orbitals can be obtained directly as a function of the signs taken on by the variables x, y and z. To do this, we use the correspondence between the spherical and Cartesian coordinates given (see Figure 7.3) by:

$$\begin{cases} \dfrac{x}{r} = \sin\theta \ \cos\varphi \\ \dfrac{y}{r} = \sin\theta \ \sin\varphi \\ \dfrac{z}{r} = \cos\theta \end{cases}$$

(for example, if $\cos\theta < 0$, $\rightarrow z < 0$ and Ψ_0; see also Figure 7.2).

With $f(r) = \frac{1}{2}\sqrt{\frac{3}{\pi}}\,\frac{R(r)}{r} > 0$ (and $r \approx r_f$), the functions Ψ_0, Ψ_1 and Ψ_1 can thus be written in the following way:

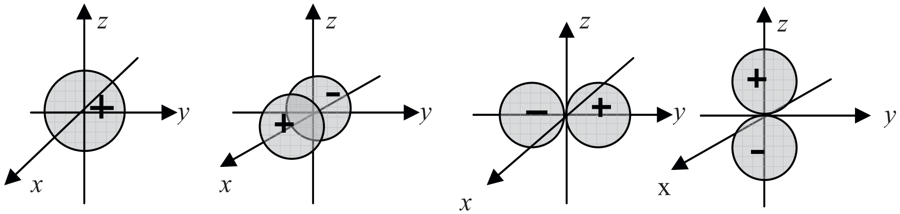
$$\Psi_1 = X = x\,f(r) = \Psi_{px} \qquad (\Psi_{px} > 0 \text{ or } < 0 \text{ depending on whether } x > 0 \text{ or } x < 0)$$

$$\Psi_{-1} = Y = y\,f(r) = \Psi_{py} \qquad (\Psi_{py} > 0 \text{ or } < 0 \text{ depending on whether } y > 0 \text{ or } y < 0)$$

$$\Psi_0 = Z = z\,f(r) = \Psi_{pz} \qquad (\Psi_{pz} > 0 \text{ or } < 0 \text{ depending on whether } z > 0 \text{ or } z < 0)$$

Finally, the shape of the *s*- and *p*-orbitals can be schematically illustrated as in Figure 7.4 (r is fixed to a value denoted r_f which is such that $\int_0^{r_f} r^2 \left|R(r)\right|^2 dr =$ 95%).

To conclude, we can state that the *s*-orbitals are non-directional, as opposed to the *p*-orbitals that are.



Orbitals: *ns* *np$_x$* *np$_y$* *np$_z$*

nodal plane: (yOz) (xOz) (xOy)
(nodal plane is where there is zero probability of presence)

Figure 7.4. *Representations of ns and np atomic orbitals*

NOTE.– The orbitals represented above are for one electron; these are the orbitals that will be used when building molecular orbitals when there is one electron that will be shared with a neighboring atom, as in a covalent bond. The atomic orbitals for an atom with i electrons are given by $\Psi = \Pi_i\,\Psi_i$ where the Ψ_i are wave functions for each i electron. The former Ψ are asymmetric (Pauli) and are placed under a Slater determinate form.

7.1.2. *Molecular orbitals*

7.1.2.1. *Classic case of the H_2^+ molecular ion*

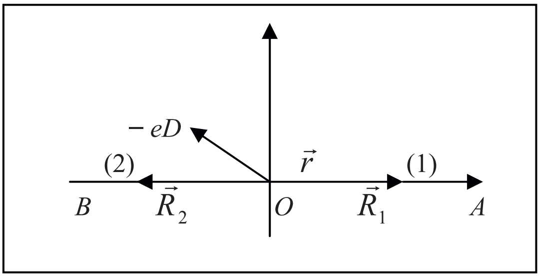


Figure 7.5. *Layout of ions and electron in H_2^+*

A simple example is that of the first approximation where the overlap of adjacent orbitals is neglected (otherwise the supplementary condition $\partial E/\partial k = O$ should be used where there is a minimization of the energy according to a method of variation). With the two nuclei (protons) being situated at $\vec{R}_1$ and $\vec{R}_2$ and with the electron common to $\vec{r}$ (see Figure 7.5), the Hamiltonian of the system is (assuming that the nuclei are fixed in an adiabatic approximation):

$$H = -\frac{\hbar^2}{2m}\Delta - \frac{e^2}{4\pi\varepsilon_0\left(\left|\vec{r}-\vec{R}_1\right|\right)} - \frac{e^2}{4\pi\varepsilon_0\left(\left|\vec{r}-\vec{R}_2\right|\right)}$$

If $\varphi(r) = A\exp(-r/a_1)$ is the wave function of the fundamental state of the hydrogen atom, this function is still a solution to the problem for (H_2^+) if we assume that the electron is localized preferentially on one of the two nuclei:

– if the electron is localized about (1), $\varphi(r) \to \varphi\left(\left|\overrightarrow{AD}\right|\right) = \varphi\left(\left|\vec{r}-\vec{R}_1\right|\right) = \varphi_1$;

– if the electron is localized about (2), $\varphi(r) \to \varphi\left(\left|\overrightarrow{BD}\right|\right) = \varphi\left(\left|\vec{r}-\vec{R}_2\right|\right) = \varphi_2$.

If the electron is localized between the two nuclei, the solution must correspond to a mixture of the two preceding states. This implies a search for a wave function φ for the molecular ion H_2^+ with a wave function in the form $\varphi = c_1\,\varphi_1 + c_2\,\varphi_2$. If the electron is localized at (1) then $c_1 \to 1$ and $c_2 = 0$; if the electron is localized at (2), then $c_1 = 0$ and $c_2 \to 1$.

This function must satisfy the equation for proper values:

$$H \mid \varphi > = E \mid \varphi > \quad \text{so that} \quad H \mid c_1\, \varphi_1 + c_2\, \varphi_2 > = E \mid c_1\, \varphi_1 + c_2\, \varphi_2 >$$

On multiplying the left-hand side of the equation by $< \varphi_1 \mid$ we obtain:

$$c_1 < \varphi_1 \,|\mathrm{H}|\, \varphi_1 > + \, c_2 < \varphi_1 \,|H|\, \varphi_2 > = E\, c_1 < \varphi_1 \mid \varphi_1 > + \, \mathrm{E}\, c_2 < \varphi_1 \mid \varphi_2 >$$

With the simplifying hypothesis which makes it possible to neglect the overlap integral $S = < \varphi_1 \mid \varphi_2 > = 0$, and with making $H_{11} = < \varphi_1 \,|H|\, \varphi_1 >$ and $H_{12} = < \varphi_1 \,|H|\, \varphi_2 >$, we find that:

$$c_1\, H_{11} + c_2\, H_{12} = E\, c_1$$

Similarly, multiplying the left-hand side with $< \varphi_2 \,|$, and with $H_{ij} = < \varphi_i \,|H|\, \varphi_j >$, we have:

$$c_1\, H_{21} + c_2\, H_{22} = E\, c_2$$

We thus obtain a system with two equations and two unknowns (c_1 and c_2), that can now be written as:

$$\begin{cases} c_1\,(H_{11} - E) + c_2\, H_{12} = 0 & [7.4] \\ c_1\, H_{21} + c_2\,(H_{22} - E) = 0 & [7.5] \end{cases}$$

In addition,

– the wave functions φ_1 and φ_2 are real, thus making it possible to state that:

$$H_{12} = < \varphi_1 \,|H|\, \varphi_2 > = < \varphi_2 \,|H|\, \varphi_1 > = H_{21} = -\,\beta \text{ (with } \beta > 0\text{);}$$

– also, the problem (and hence the Hamiltonian) remains invariant with respect to any permutation of the nuclei A (index 1) and B (index 2), such that $H_{11} = H_{22} = -\alpha$ (when $\alpha > 0$).

The preceding system of equations [7.4]. [7.5] with two unknowns finally yields a non-trivial solution when its determinate is zero. Using our notation, we have the equation $(\alpha + E)^2 - \beta^2 = 0$, which gives two solutions $(\alpha + E) = \pm\beta$ for energy that can be written as:

– $E_L = -\alpha - \beta$ (the lowest energy corresponding to the bonding level as α and β are positive); and

– $E_A = -\alpha + \beta$ (the highest and therefore most unstable energy which corresponds to an anti-bonding level). (It can also be noted along the way that $E_A - E_L = 2\beta$.)

Placing E_A into equations [7.4] and [7.5] gives us $(c_1 + c_2)\beta = 0$, so that $c_1 = -c_2 = c_A$. Similarly, by moving E_L into [7.4] and [7.5], we obtain $(-c_1 + c_2)\beta = 0$, so that $c_1 = c_2 = c_L$.

The wave function for the H_2^+ system thus accepts two solutions and they can be written as:

$$\Psi_L = c_L(\varphi_1 + \varphi_2)$$
$$\Psi_A = c_A(\varphi_1 - \varphi_2)$$

The normalization condition for the two functions Ψ_L and Ψ_A:

$$< \Psi_L | \Psi_L > = < \Psi_A | \Psi_A > = 1$$

makes it possible to determine that $c_L = c_A = (1/\sqrt{2})$ and that:

$$\Psi_L = \frac{1}{\sqrt{2}}(\varphi_1 + \varphi_2) \qquad [7.6]$$

$$\Psi_A = \frac{1}{\sqrt{2}}(\varphi_1 - \varphi_2) \qquad [7.7]$$

We can thus go onto schematically represent these wave functions. Figure 7.6 shows the bonding solution where the electron has a high probability of presence between the two nuclei, so that the strong attraction between the electron and the nuclei bond the system. However, for the anti-bonding solution, the electron has a strong probability of presence on either outer side of the two nuclei such that the strong electrostatic repulsion between nuclei (1) and (2) destabilize the system.

NOTE.– With $S = < \varphi_1 \mid \varphi_2 > \neq 0$ (overlap integral), the same but longer calculation gives:

$$\begin{aligned} \Psi_L &= \frac{\varphi_1 + \varphi_2}{\sqrt{2(1+S)}} \qquad & E_L &= -\frac{\alpha + \beta}{1+S} \\ \Psi_A &= \frac{\varphi_1 - \varphi_2}{\sqrt{2(1-S)}} \qquad & E_A &= -\ \frac{\alpha - \beta}{1-S} \end{aligned} \tag{7.8}$$

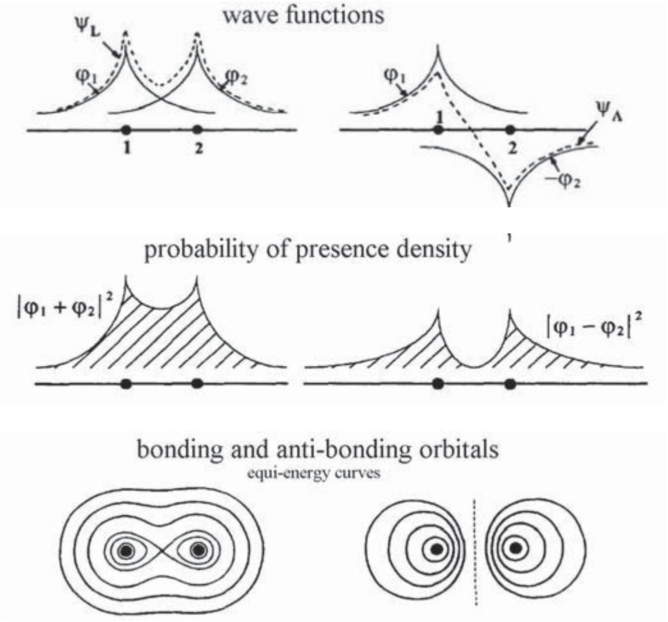


Figure 7.6. *Wave functions and orbitals for H_2^+ (bonding state on the left and anti-bonding state on the left)*

7.1.2.2. *The H_2 molecule*

In this system, there are two electrons and they are assumed to be at the same potential as the electron in the H_2^+ system. This is because to a first approximation it is possible to assume that the interaction potential between the two electrons is negligible.

Each electron has orbital states Ψ_L and Ψ_A, the lowest energy corresponding to the two electrons both being in a bonding orbital. In this state the electrons are localized between the two nuclei (where their probability of presence is highest) and the energy of the system is minimized because the electrons benefit from the attractive potential (negative potential energy) generated by the overlapping Coulombic potentials of the positive nuclei. This bonding charge, localized between the two nuclei (indicated by the maximum electronic density), is the basis for the concept of covalence.

The final outcome is as if we had melted the states of each hydrogen into one another so as to increase the local electronic density between the two nuclei. The bonding state is again obtained with the help of a linear additive combination of the two individual orbitals of the atoms. This can be schematically illustrated as in Figure 7.7.

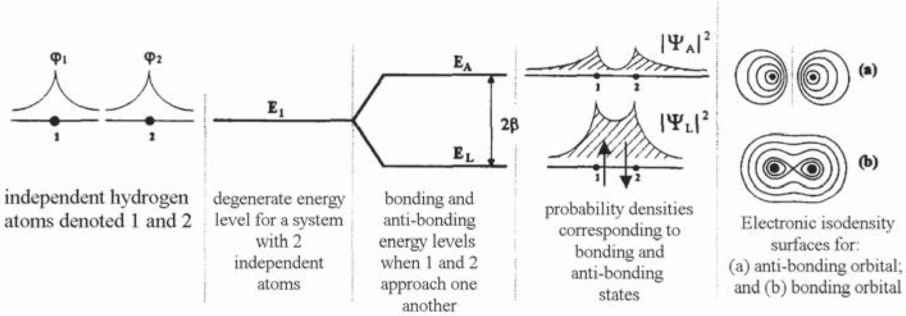


Figure 7.7. *Energy and electron states of* H_2

The wave function written in general terms for bonding or anti-bonding states as $\Psi = c_1\,\varphi_1 + c_2\,\varphi_2$ appears as a linear combination of atomic orbitals (LCAO) and is a well known method for finding molecular orbitals.

NOTE.– The Ψ_L function is symmetric while Ψ_A is asymmetric. Given that there are two electrons to position for the example of molecular hydrogen, we find that:

– on the E_L level, the spins of the two electrons must be anti-parallel in order to be distinguished and thus $S = 0$ (singlet state);

– on the E_A level, the spins of the two electrons can be parallel as they are already differentiated by their orbital levels (Ψ_A changes sign if the electrons are permuted). In this case, S can equal both 0 or 1 (i.e. singlet or triplet states).

7.1.3. *σ- and π-bonds*

7.1.3.1. *σ-bonds*

These are molecular orbitals which have a symmetry about an axis drawn between the two atoms that are covalently bonded. From *s*-orbitals (the fundamental state for hydrogen) we have as an example the structure shown in Figure 7.8.

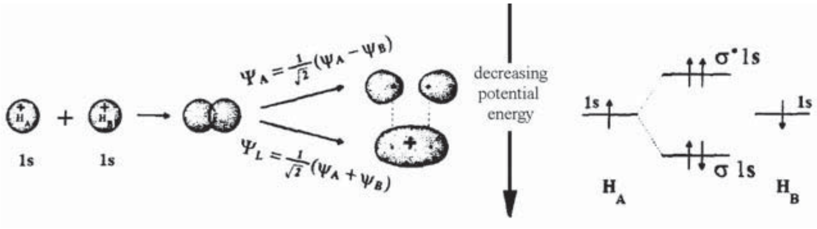


Figure 7.8. *σ 1s and σ* 1 s-orbitals*

If the axis (*Ox*) is that on which the two atoms *A* and *B* are placed then the $2p_x$ orbitals will have an appearance much like that shown in Figure 7.9.

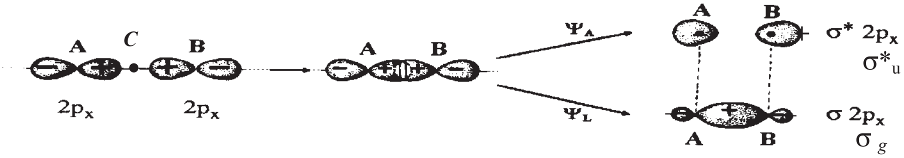


Figure 7.9. *σ 2p and σ* 2p orbitals*

where:

$$\Psi_L = \frac{1}{\sqrt{2}}(\Psi_{A2p_x} + \Psi_{B2p_x}) \text{ and } \Psi_A = \frac{1}{\sqrt{2}}(\Psi_{A2p_x} - \Psi_{B2p_x}).$$

7.1.3.2. *π-bonds*

These are molecular orbitals that take on a plane of symmetry than goes through the line O*x* that joins the atoms and is perpendicular to the direction O*y* for $\pi 2p_z$ orbitals or perpendicular to the direction O*z* for $\pi 2p_y$ orbitals.

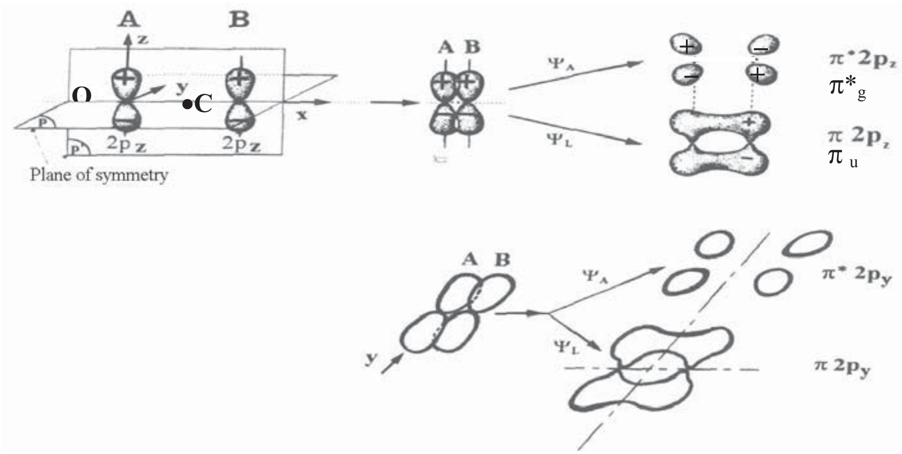


Figure 7.10. *$\pi\,2p_z$, $\pi^*\,2p_z$, $\pi\,2p_y$, and $\pi^*\,2p_y$ orbitals*

NOTE.– The molecular orbitals are termed paired (and denoted g for “gerade” from the German) and when following a symmetric operation around the center (*C*) of symmetry of the molecule the sign of the orbital does not change. When the sign does change, the orbitals are denoted by with the letter *u* to indicate “ungerade”. So, bonding σ-orbitals are paired and denoted σ_g, while anti-bonding orbitals σ* are unpaired and denoted σ^*_u. Bonding π-orbitals are unpaired and denoted π_u, while anti-bonding π*-orbitals are denoted π^*_g.

7.1.4. *Conclusion*

Following the introduction of the principal types of atomic and molecular orbitals (bonding across two atoms) we will now look at an assembly of bonds in a linear atomic chain (1D solid).

7.2. Form of the wave function in strong bonds: Floquet’s theorem

7.2.1. *Form of the resulting potential*

An infinite chain of atoms each spaced a distance *a* apart (thus *a* is the lattice period) is shown in Figure 7.11. The distance *a* is such that $a \geq 2R$ where *R* is defined in the figure. The atoms in the chain are numbered …(0), (1), (2), …(*s*–1), (*s*),…, so that by placing the atom denoted (0) at the origin, any atom called *s* has a distance from the origin given by $r_s = sa$.

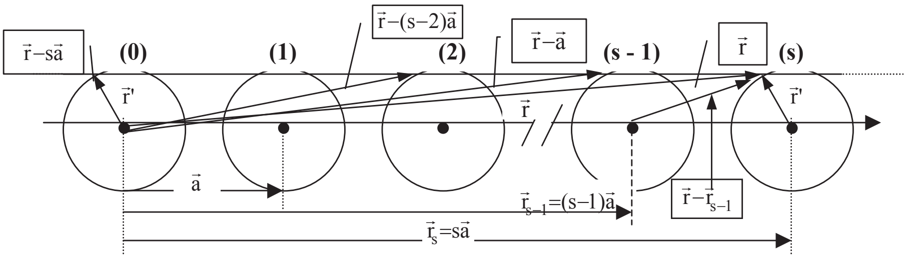


Figure 7.11. *Layout of atoms in an infinite 1D chain*

So within this geometric system, the potential generated by an atom (*s*) at a point in space located by $\vec{r}$, such that $v_s(\vec{r})$, is the same as that generated by an atom (0) at a point $\vec{r}\,'$ and similarly denoted $v_0(\vec{r}\,')$. By making $\overrightarrow{r_s} = s\vec{a}$, and on remarking that $\vec{r}_s + \vec{r}\,' = \vec{r}$, we finally have $v_s(\vec{r}) = v_0(\vec{r}\,') = v_0(\vec{r} - \vec{r}_s)$.

Outside of the action of an atom (*s*), an electron placed at $\vec{r}$ will be subject to the action of neighboring atoms, with the atom (*s* – 1) generating at $\vec{r}$ the potential $v_{s-1}(\vec{r}) = v_0(\vec{r} - \vec{r}_{s-1})$, and so on.

The resulting potential at $\vec{r}$ can thus be defined for a chain of infinite length by:

$$V(\vec{r}) = \sum_{s=-\infty}^{s=+\infty} v_s(\vec{r}) = \sum_{s=-\infty}^{s=+\infty} v_0(\vec{r} - \vec{r}_s) \qquad [7.9]$$

Now we can state that the resultant potential $V(\vec{r})$ is periodic with a period of $\vec{a}$, as the symmetry of the problem of an infinite chain demands that the resultant potential at $\vec{r}$ is the same as that calculated at $\vec{r} - \vec{a}$, …, $\vec{r} - s\vec{a}$, and so on whether the whole number *s* be negative or positive. This means that:

$$V(\vec{r}) = V(\vec{r} - \vec{a}) = V(\vec{r} - s\vec{a}) \qquad [7.10]$$

Each electron is dropped into a potential $V(\vec{r})$ that displays a periodic function, and therefore the wave function is a Bloch function. This result will be used in section 7.2.3 to establish Floquet's theorem.

7.2.2. *Form of the wave function*

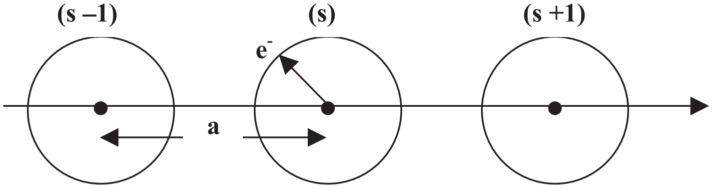


Figure 7.12. *A system of practically independent atoms*

If the atoms are well spaced out from one another, as in Figure 7.12, the valence electron attached to each nucleus *s* is not affected by the nucleus's neighbors. And in this case the potential discerned by an electron placed at $\vec{r}$ can be reduced to:

$$v_s(\vec{r}\,) = v_0\left(\vec{r} - \vec{r}_s\right)$$

and thus we have $\mathrm{V}(\vec{r}) \approx v_0\left(\vec{r} - \vec{r}_s\right)$.

Similarly, the wave function for the electron is that obtained for a single atom *s*, that is an atomic wave function [$\psi_s(\vec{r}\,)$].

If the atoms are brought closer to one another, then neighbors will start to provoke effects on the electron placed at $\vec{r}$, and the form of the wave function will resemble that given for a molecular orbital by the LCAO method (see section 7.1.2), as in:

$$\psi(\vec{r}) = \sum_s c_s \psi_s(\vec{r}) \qquad [7.11]$$

where $\psi_s(\vec{r})$ is the wave function of an electron placed at $\vec{r}$ and belonging to an atom denoted *s*.

We can see that the form of equation [7.11] resembles that of [7.9] for the potential followed by the electron.

Using an argument similar to that applied to the potential makes it possible to state that the wave function $\psi_s(\vec{r})$ for the electron placed at $\vec{r}$ belonging to the atom *s* is identical to that of the same electron placed at $\vec{r}\,'$ on the atom 0, so that:

$$\psi_s(\vec{r}) = \psi_0(\vec{r}\,'), \text{ and then with } \vec{r} = \vec{r}_s + \vec{r}\,',\ \psi_s(\vec{r}) = \psi_0(\vec{r} - \vec{r}_s)$$

As a definition, and by analogy to the potential described in equation [7.9], we can write that:

$$\psi(\vec{r}) = \sum_{s=-\infty}^{+\infty} c_s \psi_0(\vec{r} - \vec{r}_s) \qquad [7.12]$$

7.2.3. *Effect of potential periodicity on the form of the wave function and Floquet's theorem*

We can now state that the wave function for an electron placed at $\vec{r}$ and belonging to an infinitely long chain can be written:

– either in the form of a linear combination from equation [7.12]: $\psi(\vec{r}) = \sum_{s=-\infty}^{+\infty} c_s \psi_0(\vec{r} - \vec{r}_s)$, this being a Hückel development of the wave function;

– or in the form of a Bloch function, with the potential to which the electron is submitted being periodic so that (see Chapter 3):

$$\psi_k(\vec{r}) = e^{i\vec{k}\vec{r}} u(\vec{r}), \text{ where } u(\vec{r}) = u(\vec{r} + \vec{a}) \qquad [7.13]$$

The Bloch form applied to a wave function calculated for $\vec{r} + \vec{a}$ makes it possible to state that:

$$\psi_k(\vec{r} + \vec{a}) = e^{i\vec{k}(\vec{r} + \vec{a})} u(\vec{r} + \vec{a}) = e^{i\vec{k}\vec{a}} e^{i\vec{k}\vec{r}} u(\vec{r}) = e^{i\vec{k}\vec{a}} \psi_k(\vec{r}) \qquad [7.14]$$

and, by developing $\psi_k(\vec{r})$ as found in equation [7.14] and according to Hückel from equation [7.12], we find an initial expression for $\psi_k(\vec{r} + \vec{a})$:

$$\begin{aligned}\psi_k(\vec{r}+\vec{a}) = e^{i\vec{k}\vec{a}}[\ldots + c_0\psi_0(\vec{r}) + c_1\psi_0(\vec{r}-\vec{a}) \\ + \ldots + c_s\psi_0(\vec{r}-s\vec{a}) + \ldots]\end{aligned} \quad [7.15]$$

However, the Hückel development directly carried out on $\psi_k(\vec{r}+\vec{a})$ gives:

$$\begin{aligned}\psi_k(\vec{r}+\vec{a}) = \ldots + c_0\psi_0(\vec{r}+\vec{a}) + c_1\psi_0(\vec{r}+\vec{a}-\vec{a}) + \ldots \\ + c_s\psi_0(\vec{r}+\vec{a}-s\vec{a}) + c_{s+1}\psi_0(\vec{r}+\vec{a}-[s+1]\vec{a}) + \ldots\end{aligned} \quad [7.16]$$

The identification, term by term, between equation [7.15] and [7.16] gives, for example:

$e^{i\vec{k}\vec{a}}c_0\psi_0(\vec{r}) = c_1\psi_0(\vec{r})$, so that $c_1 = c_0\ e^{i\vec{k}.\vec{a}}$.

$e^{i\vec{k}\vec{a}}\ c_1\psi_0(\vec{r}-\vec{a}) = c_2\psi_0(\vec{r}\ +\ \vec{a}\ -\ 2\vec{a})]$, so that $c_2\ = c_1 e^{i\vec{k}\vec{a}}$

$e^{i\vec{k}\vec{a}}c_s\psi_0(\vec{r}-s\vec{a})] = c_{s+1}\psi_0(\vec{r}+\vec{a}-[s+1]\vec{a})$, so that $c_{s+1}=c_s e^{i\vec{k}\vec{a}}$

In general terms we thus have:

$$c_s = c_{s-1}\ e^{i\vec{k}\vec{a}} = c_{s-2}\ e^{i\vec{k}\cdot 2\vec{a}} = \ldots = c_0 e^{i\vec{k}\cdot s\vec{a}} = c_0 e^{i\vec{k}\cdot\vec{r}_s}. \quad [7.17]$$

With this we are brought to the final form of the wave function (Floquet's theorem) that can be written in two equivalent forms (using the notation $\psi(\vec{r}) \equiv \psi_k(\vec{r})$):

$$\begin{aligned}&\psi(\vec{r}) = \sum_s c_s\psi_s(\vec{r}) \text{ so that with } c_s\ = c_0 e^{i\vec{k}\cdot\vec{r}_s} \\ &\psi(\vec{r}) = \psi_k(\vec{r}) = c_0\ \sum_s e^{i\vec{k}\cdot\vec{r}_s}\psi_s(\vec{r}),\end{aligned} \quad [7.18]$$

or rather:

$$\psi_k(\vec{r}) = \sum_{s=-\infty}^{+\infty} c_s\psi_0(\vec{r}-\vec{r}_s) = c_0\sum_{s=-\infty}^{+\infty} e^{i\vec{k}.\vec{r}_s}\psi_0(\vec{r}-\vec{r}_s) \text{ with } \vec{r}_s = s\vec{a} \quad [7.19]$$

7.3. Energy of a 1D system

7.3.1. *Mathematical resolution in 1D where x ≡ r*

For an atom *s* in a linear chain as shown in Figure 7.13, the resultant potential V(x) is that shown with a dotted line. If the atoms are far enough apart so that each atom can be considered independent, then the potential follows $U_0(x)$.

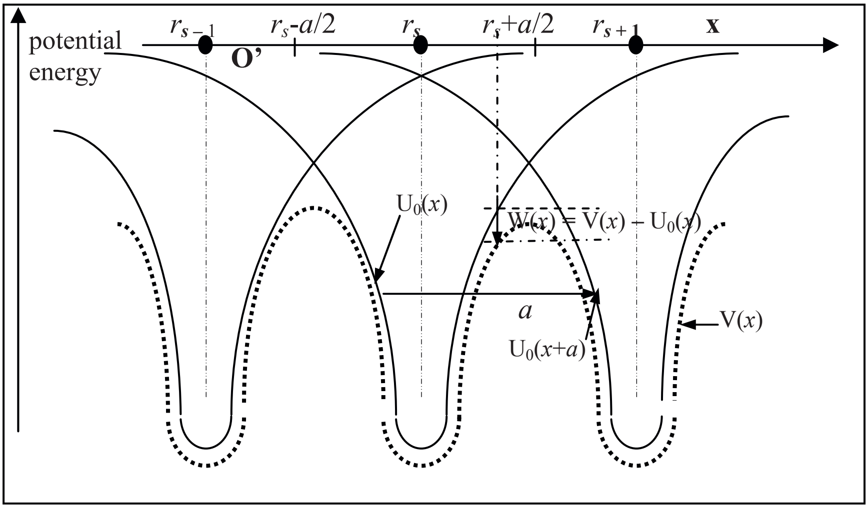


Figure 7.13. *Explanation of the W(x) = V(x) – U₀(x) function*

As shown in Figure 7.13, the potential $U_0(x)$ generated along x by the single atom *s* is the same as the potential $U_0(x + a)$ generated by the atom (s + 1) along (x + a). We thus find that $U_0(x) = U_0(x + a)$ and the potential U_0 produced by the independent atoms is periodic. We note also that for $x \in \left]r_s - \frac{a}{2}, r_s + \frac{a}{2}\right[$ we have $V(x) \approx U_0(x)$, and this becomes all the more true the closer we are to r_s (at a mid-point from $r_s - a/2$ and $r_s + a/2$). We thus make $W(x) = V(x) - U_0(x)$ where W(x) is small. In addition, as $V(x) < U_0(x)$, we also have $W(x) < 0$.

So going through the successive Schrödinger equations:

– for an electron described by the wave function $\psi_k(x)$ and placed along x so that the resulting potential is V(x) (and where $\psi_k(x)$ and V(x) take into account the effects of neighboring atoms):

$$E\psi_k(x) = -\frac{\hbar^2}{2m}\Delta\psi_k(x) + \mathrm{V}(x)\,\psi_k(x) \qquad [7.20]$$

– for an electron positioned with respect to x and belonging to an isolated atom numbered s, its wave function is $\psi_s(x) = \psi_0(x - sa)$, while the potential is $\mathrm{U}_0(x) = \mathrm{U}_0(x - sa)$. Denoting its energy with E_0, we have:

$$E_0\psi_0(x-sa) = -\frac{\hbar^2}{2m}\Delta\psi_0(x-sa) + \mathrm{U}_0(x)\,\psi_0(x-sa) \qquad [7.21]$$

Multiplying the two sides of this equation by $c_0\, e^{ikr_s}$ (where $r_s = sa$) and summing over s:

$$E_0\sum_s c_0 e^{ikr_s}\psi_0(x-r_s) = -\frac{\hbar^2}{2m}\Delta\left[\sum_s c_0 e^{ikr_s}\psi_0(x-r_s)\right] + U_0(x)\sum_s e^{ikr_s}\psi_0(x-r_s).$$

This equation can be rewritten with equation [7.19] in mind:

$$E_0\psi_k(x) = -\frac{\hbar^2}{2m}\Delta\psi_k(x) + \mathrm{U}_0(x)\,\psi_k(x) \qquad [7.22]$$

Taking the difference between equations [7.20] and [7.22], we have:

$$(E - E_0)\,\psi_k(x) = [\mathrm{V}(x) - \mathrm{U}_0(x)]\,\psi_k(x) \qquad [7.23]$$

With the potential $V(x)$ being periodic, we have $V(x) = V(x - r_s)$, such that:

$$W(x) = [V(x) - U_0(x)] = [V(x - r_s) - U_0(x - r_s)] = W(x - r_s) \qquad [7.24]$$

Note that $W(x - r_s)$ is a periodic function of a, and as such is independent of s. This parameter can therefore be included or omitted from $\sum_s$.

Finally, equation [7.23] gives, by developing $\psi_k(x)$ according to Floquet, as in equation [7.19],

$$(E - E_0)\,\psi_k(x) = c_0 \sum_s e^{ikr_s} W(x - r_s)\psi_0(x - r_s) \qquad [7.25]$$

7.3.2. *Calculation by integration of energy for a chain of N atoms*

Multiplying equation [7.25] by $\psi_k{}^*(x)$ and integrating over all N atoms numbered from 0 to $(N - 1)$ directly gives:

$$\begin{aligned}&\left(E - E_0\right)\left\langle \psi_k(x) \middle| \psi_k(x)\right\rangle \\ &\quad = \mathrm{c}_0 \sum_s e^{ikr_s} \int \psi_k^*(x)\ \mathrm{W}(x - r_s)\psi_0(x - r_s)dx.\end{aligned} \qquad [7.26]$$

With the help of equation [7.19], the calculation of $\left\langle \psi_k(x) \middle| \psi_k(x)\right\rangle$ that appears in equation [7.26] can be performed:

$$\left\langle \psi_k(x) \middle| \psi_k(x)\right\rangle = \left|c_0\right|^2 \sum_{s=0}^{s=N-1} \sum_{t=0}^{t=N-1} e^{i(ks-kt)a} \left\langle \psi_0(x - \mathrm{r_s}) \middle| \psi_0(x - \mathrm{r_t})\right\rangle$$

As $\left\langle \psi_0(x - r_s) \middle| \psi_0(x - r_t)\right\rangle = \delta_s^t$ (with $\delta_s^t = 1$ if $s = t$; $\delta_s^t = 0$ if $s \neq t$), we have:

$$\begin{aligned}\left\langle \psi_k(x) \middle| \psi_k(x)\right\rangle &= \left|c_0\right|^2 \sum_{s=0}^{s=N-1} \sum_{t=0}^{t=N-1} e^{i(ks-kt)a} \delta_s^t \\ &= \left|c_0\right|^2 \sum_{s=0}^{s=N-1} e^{i(ks-ks)a} = \left|c_0\right|^2 N.\end{aligned}$$

The normalization condition of the function $\psi_k(x)$, $<\psi_k(x)|\ \psi_k(x)> = 1 = |c_0|^2 N$, makes it possible to obtain:

$$c_0 = \frac{1}{\sqrt{N}} \qquad [7.27]$$

Equation [7.26] makes it possible to deduce (with equation [7.19]):

$$\psi_k{}^*(x) = c_0{}^* \sum_t e^{-ikr_t} \psi_0^*(x - r_t)$$

$$(E - E_0) = |c_0|^2 \sum_s e^{ikr_s} \sum_t e^{-ikr_t} \int \psi_0^*(x - r_t) \, \mathrm{W}(x - r_s) \psi_0(x - r_s) dx$$

so that with equation [7.27],

$$E - E_0 = \frac{1}{N} \sum_{s,t} e^{ik(s-t)a} \int \psi_0^*(x - ta) \, \mathrm{W}(x - sa) \psi_0(x - sa) dx$$

This expression can also be written as:

$$E = E_0 + \frac{1}{N} \sum_{s,t} e^{ik(s-t)a} \left\langle \psi_0(x - ta) \right| \mathrm{W}(x - sa) \left| \psi_0(x - sa) \right\rangle \qquad [7.28]$$

Using the Hückel conditions, which only retain couplings between first neighbors:

$$\begin{aligned} &\left\langle \psi_0(x - ta) \middle| W(x - sa) \middle| \psi_0(x - sa) \right\rangle \\ &= \left\langle \psi_t(x) \middle| W \middle| \psi_s(x) \right\rangle = \begin{cases} -\alpha \text{ when } s = t \text{ (with } \alpha > 0 \text{ as } W < 0) \\ -\beta \text{ when } s = t \pm 1 \\ 0 \text{ for other cases.} \end{cases} \end{aligned} \qquad [7.29]$$

(as $W < 0$, then $\beta > 0$ if the orbitals ψ_t and ψ_s have the same sign). In this approximation, expression [7.28] for energy following the sum over s (that varies from $s_0 = 0$ to $s_{N\text{-}1} = N - 1$) gives

$$\begin{aligned} E = E_0 + \frac{1}{N} \Bigg[&\sum_t e^{ik(s_0 - t)a} \left\langle \psi_t \middle| W \middle| \psi_{s_0} \right\rangle \\ &+ \sum_t e^{ik(s_1 - t)a} \left\langle \psi_t \middle| W \middle| \psi_{s_1} \right\rangle \\ &+ \sum_t e^{ik(s_{N-1} - t)a} \left\langle \psi_t \middle| W \middle| \psi_{s_{N-1}} \right\rangle \Bigg] \end{aligned} \qquad [7.30]$$

Each term of the sum between the brackets that carries N terms (as s has N values) in fact gives the same contribution, with each being of the form:

$$\sum_t e^{ik(s_j - t)a} \left\langle \psi_t \right| W \left| \psi_{s_j} \right\rangle$$

$$= \underbrace{e^{ik(0)\,a}\,(-\alpha)}_{s_j = t} + \underbrace{e^{ik(-1)\,a}\,(-\beta)}_{s_j = t-1} + \underbrace{e^{ik(+1)\,a}\,(-\beta)}_{s_j = t+1} = -\alpha - \beta\, e^{-ika} - \beta\, e^{ika}$$

Finally the bracket of equation [7.30] is equal to $N\,[-\alpha - \beta\, e^{-ika} - \beta\, e^{ika}]$, and the energy ($E$) for the expression is now:

$$E = E_0 - \alpha - \beta\, e^{-ika} - \beta\, e^{ika} = E_0 - \left[\alpha + 2\beta\left(\frac{e^{ika} + e^{-ika}}{2}\right)\right]$$

$$= E_0 - [\alpha + 2\,\beta \cos ka],$$

which can be rewritten:

$$E = E_0 - \alpha - 2\,\beta \cos ka \qquad [7.31]$$

The graphical representation of $E = f(k)$ (energy dispersion curve) is given in Figure 7.14. It shows that the amplitude of the variation in E as a function of k amounts to 4β; the permitted bands are as wide as β are large (strong transfer integral between electrons on closest neighbors).

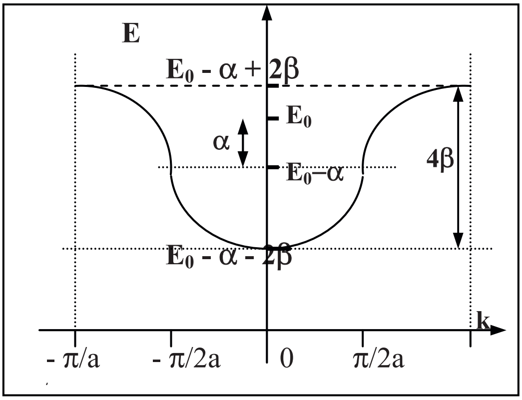


Figure 7.14. *The dispersion curve E = f(k) from strong bond and Hückel approximations*

7.3.3. *Note 1: physical significance in terms of ($E_0 - \alpha$) and β*

From equation [7.21],

$$E_0\psi_0(x-sa) = -\frac{\hbar^2}{2m}\Delta\psi_0(x-sa) + \mathrm{U}_0(x)\ \psi_0(x-sa),$$

and equations [7.29]:

$$\begin{cases} E_0 = \left\langle \psi_0(x-sa) \right| -\dfrac{\hbar^2}{2m}\Delta + \mathrm{U}_0(x) \left| \psi_0(x-sa) \right\rangle \\ -\alpha = \left\langle \psi_0(x-sa) \right| \mathrm{W}(x-sa) \left| \psi_0(x-sa) \right\rangle \text{ (one of the Hückel conditions [7.29])} \end{cases} \qquad [7.32]$$

so that with W($x - sa$) = W(x) (period of W: equation [7.24]):

$$\mathrm{E}_0 - \alpha = \left\langle \psi_0(x-sa) \right| -\frac{\hbar^2}{2m}\Delta + \mathrm{U}_0(x) + \mathrm{W}(x) \left| \psi_0(x-sa) \right\rangle$$

$$= \left\langle \psi_s(x) \right| -\frac{\hbar^2}{2m}\Delta + \mathrm{U}_0(x) + \mathrm{W}(x) \left| \psi_s(x) \right\rangle$$

so that:

$$(E_0 - \alpha) = \left\langle \psi_s(x) \right| -\frac{\hbar^2}{2m}\Delta + V(\mathrm{x}) \left| \psi_s(x) \right\rangle \qquad [7.33]$$

as V(x) = $\mathrm{U}_0(x)$ + W(x).

The bracketed terms indicate that:

– the E_0 term represents the energy of an electron situated on a given atom s within a potential generated only by that atom (potential is $\mathrm{U}_0(x)$ as detailed in equation [7.32]);

– the ($-\alpha$) term represents the energy of an electron (on a given atom s) influenced by atoms that are neighbors to the principal atom (potential is W(x) as given in equation [7.29]);

– the $(E_0 - \alpha) = E'_0$ term represents the energy of an electron situated on a given atom s and placed within the general resultant potential given by V(x). This comes from equation [7.33], where the potential V(x) = U$_0$(x) + W(x) generated by this atom gives potential U$_0$(x) and its neighboring atoms (potential W(x), can be seen in zone I of Figure 7.15.

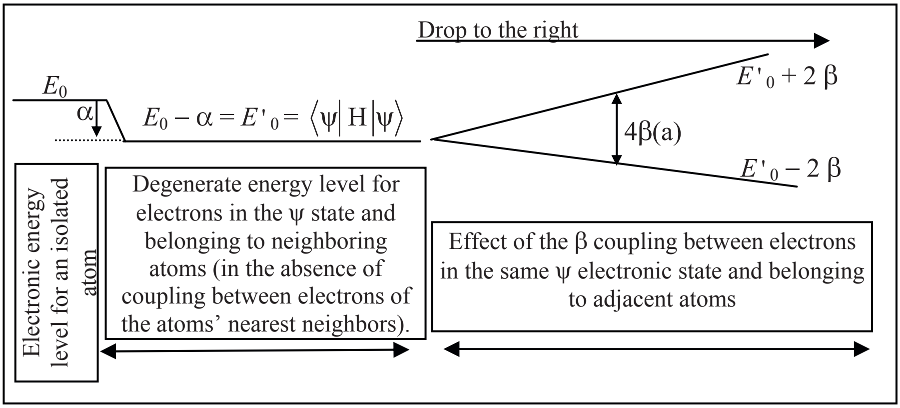


Figure 7.15. *Schematic illustration of the effect of coupling between electrons in the same electronic state (in zone II, we find that "a" decreases and the interaction between atoms and $\beta_{(a)}$ increase)*

For its part, β was defined such that (equations [7.29]):

$$\beta = \langle\psi_0(x-ta)|\ \mathrm{W}(x-sa)|\psi_0(x-sa)\rangle = \langle\psi_t|\ W|\psi_s\rangle$$

where $s = t \pm 1$. This term thus gives the coupling energy between an electron on a given atom s with electrons in the same state but belonging to t adjacent atoms (i.e. $t = s \pm 1$). The coupling is through the perturbation potential W(x), which is produced by neighboring atoms. It is this that gives rise to the level of degeneration (zone II of Figure 7.15) which corresponds to the band of permitted energies shown in Figure 7.14. This mechanism is similar to that of two interacting wells detailed in the supplementary study at the end of Chapter 1. The term thus corresponds to the bonding energy of an electron of a given atom s. As the perturbation (W) caused by adjacent atoms increases as they come closer to one another (or in other words as the lattice period a decreases), the permitted band also increases, and β = β(a) (zone II in Figure 7.15). In addition, the term −β can also be seen as the energy, of the electronic population, associated with the overlap integral, as in $S_{t,s} = \langle\psi_t|\psi_s\rangle$.

7.3.4. *Note 2: simplified calculation of the energy*

With $H = -\frac{\hbar^2}{2m}\Delta + V(x)$, according to equation [7.32] we have:

$$(E_0 - \alpha) = E'_0 = \langle \psi_s(x) | \, H | \psi_s(x) \rangle$$

Often, the term $(E_0 - \alpha) = E'_0$ is simply denoted by $-\alpha$ as it can be obtained through a more direct and simple route (that has the inconvenience of hiding several physical realities).

In effect, it is possible to state directly that:

$$E = E(k) = \frac{\langle \psi_k | H | \psi_k \rangle}{\langle \psi_k | \psi_k \rangle} = \frac{\sum_s \sum_t e^{ik(s-t)a} \langle \psi_s | H | \psi_t \rangle}{\sum_s \sum_t e^{ik(s-t)a} \langle \psi_s | \psi_t \rangle}.$$

Using the fact that $\langle \psi_s | \psi_t \rangle = \delta_{st}$ and noting that the Hückel approximations can be given by:

– $H_{ss} = \langle \psi_s | H | \psi_s \rangle = -\alpha$ = Coulomb integral = a negative constant by taking the origin of the energies as those of an electron at infinity;

– $H_{st} = \langle \psi_s | H | \psi_t \rangle = -\beta$ when $s \neq t$ and s and t are adjacent ($-\beta < 0$, is the resonance integral, also called the transfer integral, between electron s and electron t).

It is possible to directly obtain (as the number of upper and lower terms are identical as the sum making it possible to avoid the tortuous double sum):

$$E = E(k) = -\alpha - \beta\, e^{ika} - \beta\, e^{-ika} = -\alpha - 2\,\beta \cos(k \cdot a).$$

It should be noted that in this version of the Hückel approximation:

– on one hand, the term $H_{ss} = -\alpha$ is identical to that of $E'_0 = E_0 - \alpha$ from section 7.2.2; and

– on the other hand, the term $H_{st} = -\beta$ when $s = t \pm 1$, can be written as:

$$\mathrm{H_{st}} = \langle \psi_s | H | \psi_t \rangle = \langle \psi_s | -\frac{\hbar^2}{2m}\Delta + V(x) | \psi_t \rangle$$

$$= \langle \psi_s | -\frac{\hbar^2}{2m}\Delta + U_0(x) + W(x) | \psi_t \rangle$$

$$= E_0 \langle \psi_s | \psi_t \rangle + \langle \psi_s | W | \psi_t \rangle = \langle \psi_s | W | \psi_t \rangle .$$

and the value of $-\beta$ is that proposed at the start of section 7.2.2.

7.3.5. *Note 3: conditions for the appearance of permitted and forbidden bands*

So that a forbidden band appears (zone II in Figure 7.16), the rupture of the bands must come from two or more distinct levels (zone I in Figure 7.16). This can give the system summed in Figure 7.16, where there is a chain of atoms of which each has two distinct states given by ψ and φ and such that $E'^{\psi}_0 = \langle \psi | H | \psi \rangle|$ and $E'^{\varphi}_0 = \langle \varphi | H | \varphi \rangle|$. The gap E_G can therefore appear in a crystal made up of atoms that incorporate various different types of electrons (for example *s* and *p*), in a crystal made up of different type of atoms (see section 7.4.1), and indeed in a crystal with an asymmetric cell (see section 7.4.2).

In zone II of Figure 7.16, where the atoms are brought together, each degenerate level breaks down to permitted bands, with each of these at least initially being related to the starting state. When the rupture is sufficiently large, i.e. when *a* is small enough, for the states to mix, then the two permitted bands are separated by a forbidden band, this being at the point M in the figure. Figure 8.10, for carbon, shows this state of affairs more closely for 3D.

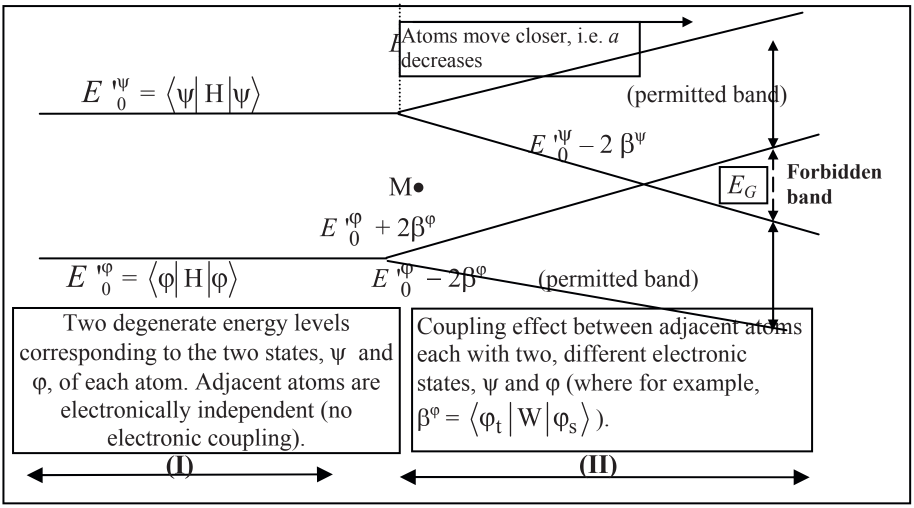


Figure 7.16. *A first approximation to the appearance of the forbidden band taking into account the electrons in the two states denoted ψ for φ of each atom*

7.4. 1D and distorted AB crystals

7.4.1. *AB crystal*

A 1D crystal made up of an alternating distribution of *A* and *B* type atoms (Figure 7.17 with in all $2n = N$ atoms) has $N/2$ *A–B* atom pairs.

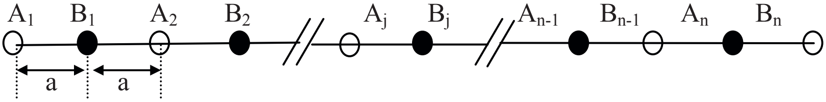


Figure 7.17. *Alternating chain of n pairs of A and B atoms, where n varies from 1 to N/2*

It is possible to see straight away that the primitive cell (which encloses one whole *AB* unit) has a period of $2a$. E_A and E_B denote the energy levels of electrons situated on atom *A* and atom *B*, respectively, and φ_A and φ_B are the corresponding wave functions.

The problem can be put into two equations similar to that of equation [7.20]. One equation relates to each atom, *A* or *B*. The states in themselves can then be sought through a linear combination of atomic orbitals for a chain of *N*/2 *A* and *B* atoms, as in:

$$\left|\psi_k\right\rangle = \sum_{\mathrm{j=1}}^{\mathrm{N/2}}\left(\nu_{\mathrm{kj}}\left|\varphi_{Aj}\right\rangle + \eta_{kj}\left|\varphi_{Bj}\right\rangle\right)$$

with, according to Floquet's theorem, $\nu_{kj} = a_k e^{ikj2a}$ and $\eta_{kj} = b_k e^{ikj2a}$. Once again using assuming poor overlap between adjacent neighbors, we have:

$$\left\langle\varphi_{Ai}\middle|\varphi_{Aj}\right\rangle = \delta_{ij} = \left\langle\varphi_{Bi}\middle|\varphi_{Bj}\right\rangle \text{ and } \left\langle\varphi_{Ai}\middle|\varphi_{Bj}\right\rangle = 0$$

and the wave function, following normalization using $\left|a_k\right|^2 + \left|b_k\right|^2 = 1$, is:

$$\left|\psi_k\right\rangle = N^{-1/2}\sum_{j=1}^{N/2}\exp\left(ikj2a\right)\left[a_k\left|\varphi_{Aj}\right\rangle + b_k\left|\varphi_{Bj}\right\rangle\right]$$

The equation for proper values is $H\left|\psi_k\right\rangle = E_k\left|\psi_k\right\rangle$, with:

$$H = -\frac{\hbar^2}{2m}\Delta + \sum_j V_{Aj} + \sum_j V_{Bj}$$

and the V_{Aj} and V_{Bj} potentials are defined as for a single type of atom.

Successive multiplication of the equation for proper values by $\left\langle\varphi_{Aj}\right|$ and then by $\left\langle\varphi_{Bj}\right|$ gives rise to two types of equations with a compatibility that is given by a second-order equation carrying the coefficients α_A and α_B of the form $\alpha_A = \langle \varphi_A | V_B | \varphi_A \rangle$ and $\alpha_B = \langle \varphi_B | V_A | \varphi_B \rangle$.

The upshot is that there are two solutions for the energy. The separation between them, $\Delta E = E^+_{\min} - E^-_{\max}$, is equal to the forbidden band. This conclusion is shown schematically in Figure 7.18. When $k = 0$, the state with the lowest energy belongs to the curve denoted E^-. It is represented by a bonding orbital, here called Ψ_L. Still with $k = 0$, the state with the highest energy is found in the curve E^+ and is represented by an anti-bonding orbital, here denoted by Ψ_A.

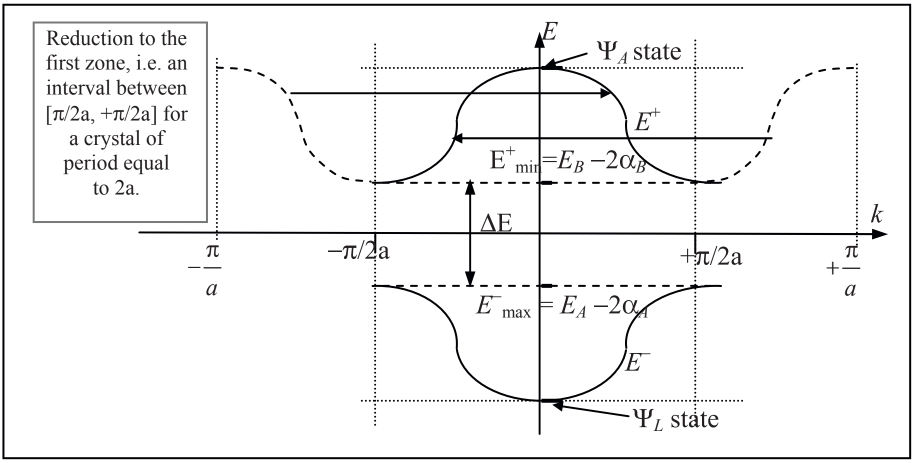


Figure 7.18. *Dispersion curve, E = f(k), for the AB diatomic system of period 2a*

7.4.2. *Distorted chain*

7.4.2.1. *Representation*

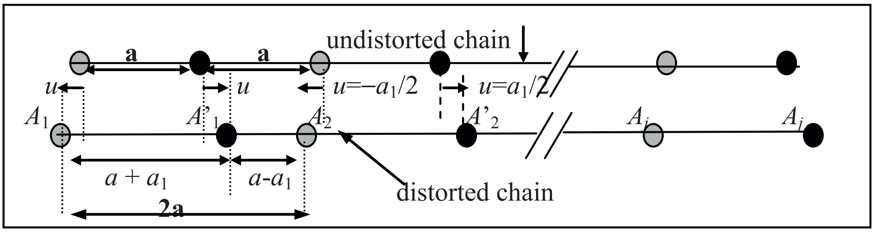
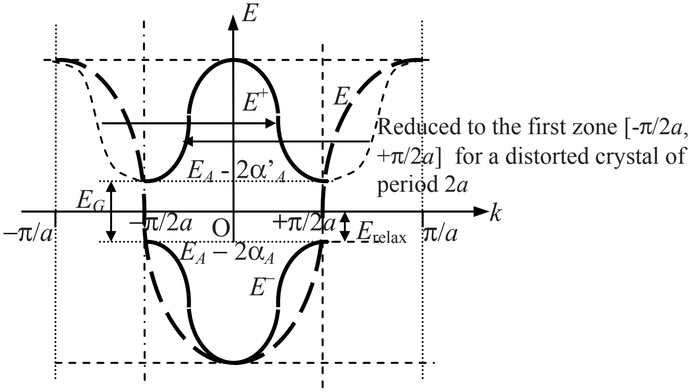


Figure 7.19. *Distorted AA' chain, where the atoms A and A' are identical but separated by bonds of alternating lengths*

If each atom (A) is displaced by a distance given by $u = \pm\,\frac{a_1}{2}$ with respect to its initial rest position in a chain with period a, then the result is the structure shown in Figure 7.19. The period is now equal to $2a$ (as in section 7.4.1), but now there is only one type of atom, i.e. $\equiv A'$), and hence $E_A = E_{A'}$.

The upshot is the band scheme shown in Figure 7.20. The $E(k)$ curve for the undistorted chain (with a period equal to a) is traced by the dashed line, while that for the distorted crystal (with a period of $2a$) is drawn as a full line. The distortion gives rise to a gap (E_G) that changes with $k = \pm\,\pi/2a$.



Reduced to the first zone [$-\pi/2a$, $+\pi/2a$] for a distorted crystal of period $2a$

Figure 7.20. *$E = f(k)$ dispersion curves for distorted (continuous line) and undistorted dashed line) crystals. Making $\left\langle \varphi A'_{j-1} \middle| H \middle| \varphi A_j \right\rangle = -\beta_1$ and $\left\langle \varphi A_j \middle| H \middle| \varphi A'_j \right\rangle = -\beta_2$, we have: $\Delta E_{(k\,=\,\pi/2a)} = E_G = 2\,(\beta_1 - \beta_2)$. In addition, supposing that β_1 and β_2 are of the form $-\,\beta_1 = -\,\beta + \alpha u$ and $-\,\beta_2 = -\,\beta - \alpha u$ (where $\alpha > 0$), calculation gives $E_G \approx 4\alpha|u|$.*

7.4.2.2. *Conditions to generate a distorted chain*

Whether or not a distorted chain will arise is governed by the energy required to produce a chain deformation (bond alternation) and the electronic energy gained through relaxation, respectively denoted E_{defor} and E_{relax}, following the opening of a gap:

– E_{defor} is the energetic cost of going from a chain of period a to a structure of period $2a$, which alternates "long" and "short" bonds, as in $(a + a_1)$ – short bond $(a - a_1)$. It can be stated as $E_{\text{defor}} = \frac{1}{2} N k_e \left|2u\right|^2 = 2k_e N u^2$, where N is the number of atoms of type A or A', $|2u|$ is the stretching or contraction modulus of the spring that mechanically ties two adjacent atoms. This brings in the harmonic coupling of atoms through an elastic force (f_e) with a constant (k_e) where $f_e = - k_e\,(2u)$. If $2u > 0$, $f_e < 0$ and is a force of attraction.

– The electronic relaxation energy (E_{relax}) is the drop in energy of the system in going from the filling with electrons of the system with period *a* to the filling of the system based on the period 2*a*. See Figure 7.20, where E_{relax} is only an approximate value corresponding to the drop in energy of electrons that may participate in the conduction of the most energetic electrons. See the notes in section 7.5.3 for a more rigorous treatment.

If $\Delta E = E_{\text{defor}} - E_{\text{relax}} < 0$, that is to say that the gain in energy during the relaxation is greater than that of the deformation, then there is a drop in the system's energy. This stabilizes to a point of deformation with a value (u_0) such that $\left[\frac{\partial(\Delta E)}{\partial u}\right]_{u=u_0} = 0$. So, the dimerization of the system, that is the generation of alternating short and long bonds, leads to an opening in the gap that changes with $k = \pm\pi/2a$. This is the so-called Peierls transition, which corresponds to a metal–insulator transition.

7.5. State density function and applications: the Peierls metal-insulator transition

7.5.1. *Determination of the state density functions*

First, a quick recall of the evaluation of the *n*(*k*) function (see also Chapters 2 and 5). In reciprocal space, the electrons are spread throughout cells of size $\Delta k = \frac{2\pi}{L}$. This cell dimension is the quantification of the *k* space and is obtained through the progressive boundary condition (PBC, or Born-von Karman conditions). At the interval $\Delta k = \frac{2\pi}{L}$, it corresponds to an average of the electronic function denoted $\psi_k(x)$, which in turn can give two states if spin is taken into account. This also makes it possible to obtain, for the same *k* state, the two functions $\psi_k^+(x)$ and $\psi_k^-(x)$. It is therefore possible to state that:

$$n(k).\ \Delta k = n(k).\left(\frac{2\pi}{L}\right)_{L=1} = 2 \Rightarrow n(k) = \frac{1}{\pi}$$

The energy state density function [$Z(E)$] is now related to $n(k)$ by $Z(E)\,dE = 2\,n(k)dk$, from which we deduce that:

$$Z(E) = \frac{2}{\pi}\frac{1}{\frac{dE}{dk}}$$

Using this with a chain of N atoms, we have equation [7.31], where $E = E_0 - \alpha - 2\beta \cos ka$, and then as $\frac{dE}{dk} = 2\beta\ a \sin ka$, it is possible to state that

$$Z(E) = \frac{1}{\pi\beta a}\frac{1}{\sin\ ka}$$

With:

$$\sin^2(ka) = 1 - \cos^2(ka) = 1 - \left(\frac{E - E_0\ +\ \alpha}{2\beta}\right)^2$$

we obtain:

$$Z(E) = \frac{1}{\pi\beta a}\cdot\frac{1}{\sqrt{1-\left(\frac{E\ +\ \alpha - E_0}{2\beta}\right)^2}}$$

For a chain of length (L) given by $L = Na$, the number of states is:

$$N(E)\ =\ Na\,Z(E)\ =\ \frac{N}{\pi\beta}\cdot\frac{1}{\sqrt{1-\left(\frac{E\ +\ \alpha-E_0}{2\beta}\right)^2}}$$

$$= \frac{2N}{\pi}\cdot\frac{1}{\sqrt{(2\beta)^2 - (E+\alpha-E_0)^2}}.$$

The graphical representation is given in Figure 7.21b.

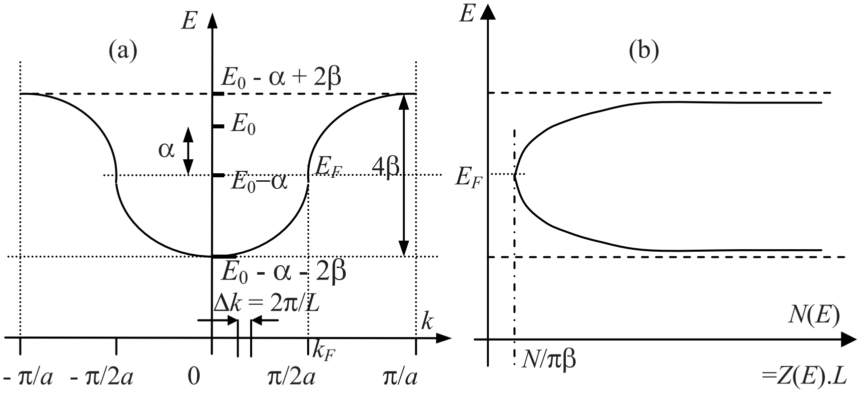


Figure 7.21. *(a) E = f(k) dispersion curve; and (b) state density curve for a chain of N atoms*

7.5.2. *Zone filling and the Peierls metal–insulator transition*

7.5.2.1. *Filling an undistorted zone*

For an undistorted chain with N atoms (length equal to Na) and N electrons (one electron per atom in a single quantum state), the number of primitive cells of dimensions $2\pi/L = 2\pi/Na$ that can be placed in a Brillouin zone of size $2\pi/a$ (placed between $k = -\pi/a$ and $k = \pi/a$) is given by:

$$\frac{\frac{2\pi}{a}}{\frac{2\pi}{Na}} = N$$

As spin makes it possible to place two electrons into each primitive cell, we can place, altogether, $2N$ electrons into the zone. Given that the system only liberates N electrons, this zone (also described by the dashed line in Figure 7.20) is only half full. The electrons can therefore easily move about as there are numerous available places, and the system is metallic.

7.5.2.2. *The metal–insulator transition*

When the distortion is energetically favored, as in the distorted crystal discussed in section 7.4.2 that gave the continuous line for $E = f(k)$ in Figure 7.20, the zone has the dimension π/a (as it is only between $k = -\pi/2a$ and $k = \pi/2a$), and only $N/2$ cells can be placed, that is only N electrons. The zone is totally filled by the system's N electrons, which can no longer move (now placed on the inside of the $E^{-} = f(k)$ line shown in Figure 7.20). A transition from a metallic to an insulating state has occurred; in other words the system has undergone a Peierls transition.

7.5.2.3. *Wave vector at the Fermi level: the position of E_F*

For a linear chain of N atoms, each liberating a single electron from the state under consideration, that has a period a and hence length $L = Na$, we can evaluate the wave vector (k_F) at absolute zero for electrons that have the highest energy, i.e. the Fermi energy which separates the occupied states from the empty states at $T = 0\ K$. At absolute zero, the temperature at which the Fermi-Dirac distribution function $F(E) = 1$ when $E < E_F$ (and $F(E) = 0$ when $E > E_F$), the number of N electrons in the chain of length L can be calculated with the help of the state density functions $Z(E)$ or $n(k)$ and must be such that:

$$N = \int_{E_{\min}}^{+\infty} F(E)N(E)dE = \int_{E_{\min}}^{E_F} N(E)dE = \int_{E_{\min}}^{E_F} L.Z(E)dE$$

$$N = \int_{-\infty}^{+\infty} N(k)F(k)dk = \int_{-k_F}^{+k_F} N(k)dk$$

$$= \int_{-k_\mathrm{F}}^{+k_\mathrm{F}} L \cdot n(k) \cdot dk = \int_{-k_\mathrm{F}}^{+k_\mathrm{F}} Na \cdot \frac{1}{\pi} dk = \frac{Na}{\pi} \cdot (2k_\mathrm{F})$$

from which:

$$k_\mathrm{F} = \frac{\pi}{2a}$$

where $N(E)$ is the state density for the chain of length L. We can thus see in Figure 7.21 that for an undistorted chain, the energy E_F that corresponds to $k_F = \frac{\pi}{2a}$ is such that $E_F = E_0 - \alpha$. For a distorted chain, that contains N atoms liberating N electrons, we again have $k_F = \frac{\pi}{2a}$. The Fermi level, at $0K$ situated midway between the occupied and empty levels, is in the middle of the gap E_G shown in Figure 7.20.

7.5.3. *Principle of the calculation of E_{relax} (for a distorted chain)*

Section 7.4.2.2 detailed how the relaxation energy is the difference between the energy of electrons belonging to an undistorted chain (E_{nd}) and those same electrons in a distorted chain, i.e. $E_{\text{relax}} = E_{\text{nd}} - E_{\text{d}}$.

The estimation of E_{relax} shown in Figure 7.20 uses the knowledge that:

– for the undistorted chain, (dashed line for $E = f(k)$), the electrons participating in the conduction are the most energetic and hence situated at E_F. As we have just seen, this energy varies with $k = k_F = \frac{\pi}{2a}$, and as shown in the figure, corresponds to the energy E_F at the intersection of the energy and $0K$ axes;

– for the distorted chain, the zone between $-\frac{\pi}{2a}$ and $\frac{\pi}{2a}$ is just filled, so that the most energetic electrons, which participate in transport, are at the summit of the band denoted E^-, that is at $E_A - 2\alpha_A$.

The energy E_{relax} shown in Figure 7.20 thus corresponds to the difference in energy of the transport electrons in the undistorted and distorted (or so-called dimerized in chemistry) chains.

A more rigorous estimation of $E_{\text{relax}} = E_{\text{nd}} - E_{\text{d}}$ can now be carried out knowing that in both cases (calculating E_{nd} and E_{d}) the electrons fill cells between $-k_F$ and $+k_F$ in the k space.

We thus have:

$$E_{nd} = \int_{-k_F}^{+k_F} E(k)N(k)dk = \int_{-\pi/2a}^{+\pi/2a} E(k)\frac{L}{\pi}dk, \text{ with } E(k) = E_0 - \alpha - 2\,\beta \cos ka$$

($E(k)$ for the undistorted chain).

Similarly, the energy (E_d) of the distorted system is given by $E_d = \int_{-\pi/2a}^{+\pi/2a} E^-(k)\frac{L}{\pi}dk$, where $E^-(k)$ is the energy function of the distorted system traced in Figure 7.20. The resulting calculations are rather long, and can be carried out as an exercise!

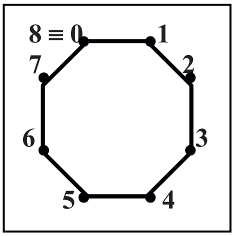


Figure 7.22. *A cyclic "linear" chain with N = 8*

7.6. Practical example of a periodic atomic chain: concrete calculations of wave functions, energy levels, state density functions and band filling

This section looks at the electronic properties of a 1D (along x) atomic chain of length L and made up of N atoms regularly spaced apart by a periodic distance called a, so that $L = Na$. A concrete treatment will be made of a closed, cyclic chain using $N = 8$.

7.6.1. *Range of variation in k*

The range is obtained from the progressive boundary conditions (PBC, or Born-von Karman conditions). As we have seen, these conditions indicate that when the chain is turned back on itself, the presence probability for an electron at a certain x coordinate is unique and does not depend on the number of turns (of length L) carried out by the electron on the chain and hence $\psi(x) = \psi(x+L)$.

The PBC conditions applied to the Bloch function (section 5.8.2.1.1) yield $e^{ikL} = 1$.

As in general terms, $1 = e^{i2p\pi}$ (with p being whole) and here with $L = Na$, we can deduce that:

$$k = k_p = 2\pi \frac{p}{Na}$$

(where p is a positive or negative integer or zero, as in $p = 0, \pm 1, \pm 2, \pm 3, \ldots$).

In addition, the energy is given by equation [7.31], so that:

$$E = E_0 - \alpha - 2\,\beta \cos ka = E_0 - \alpha - 2\,\beta \cos k_p a$$

With the $E(k)$ curve being periodic, we can limit the representation to one period (all other periods will give the same solution for energy). This gives $E(k)$ over a reduced zone that extends from $-\pi \leq ka \leq \pi$, and such that with $k = k_p$ we have:

$$-\frac{\pi}{a} \leq k_p \leq \frac{\pi}{a}$$

(reduction to the first Brillouin zone). As $k_p = 2\pi \frac{p}{Na}$, we also have $-\frac{N}{2} \leq p \leq \frac{N}{2}$, so that p takes on N successive values.

7.6.2. *Representation of energy and state density function for N = 8*

As we have seen, the domain in which $k_p = 2\pi \frac{p}{Na}$ varies can be reduced to:

$$-\frac{\pi}{a} \leq k_p \leq \frac{\pi}{a}$$

When $N = 8$, and $k_p = \frac{\pi}{4a} p$, the successive values of k_p (where $-\frac{N}{2} \leq p \leq \frac{N}{2}$, so that $p \in [-4,\ -3,..0,...3,\ 4]$) are therefore:

$$k_p = -\frac{\pi}{a},\ -\frac{3\pi}{4a},\ -\frac{\pi}{2a},\ -\frac{\pi}{4a},\ 0,\ \frac{\pi}{4a},\ \frac{\pi}{2a},\ \frac{3\pi}{4a},\ \frac{\pi}{a}.$$

Equation [7.31] for energy can thus be written as:

$$E = E_0 - \alpha - 2\,\beta \cos k_p a = E_0 - \alpha - 2\,\beta\,\cos\frac{\pi}{4} p$$

$E = f(k_p)$ and $Z(E)$ are represented in Figures 2.23a and b.

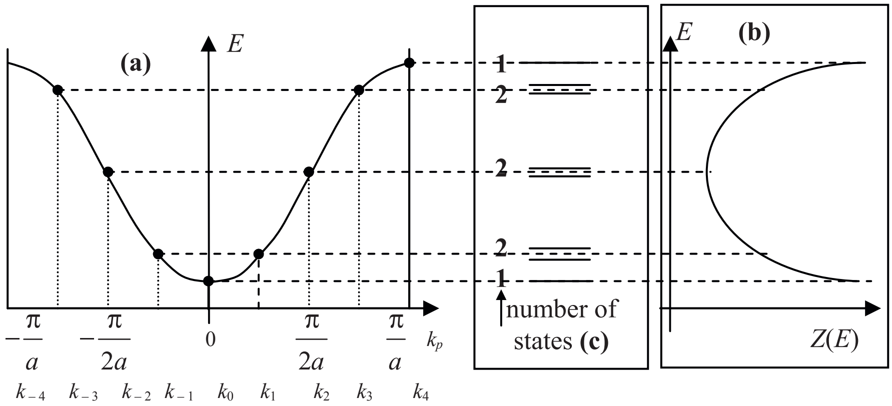


Figure 7.23. *(a) $E = f(k_p)$; (b) $Z(E)$; and (c) the number of states, all for $N = 8$*

With respect to $k_{-4} = -\frac{\pi}{a}$ and $k_4 = \frac{\pi}{a}$ the states are shared with adjacent zones, such that only one of these states is compatible.

It should be mentioned that the states at the top (k_{-3} , k_3 and k_4) and the bottom (k_{-1}, k_1 and k_0) of the band are "squashed" and in contrast to those at the middle (k_{-2} and k_2) of the band. This arises from the cosine shape of the energy curve that is flattened towards the top and bottom, and more vertically inclined near the middle. This effect increases with N, but to show this in Figure 7.23 would end up with an overload! This is because as N increases and along with it the number of values that k_p takes on, the energy levels will become extremely close to one another at the bottom and top of the band. Qualitatively, this explains the shape of $Z(E)$. In addition, this function gives the peaks at the band limits as the $E = f(k)$ curve gives a horizontal tangent to these limits. The result is that the state densities are high at the summit and bottom of the band and low in the middle (Figure 7.23b). It can also be said that if a higher value of N were given, the number of functions to trace would increase along with the complexity of the representations, without necessarily showing any more clearly what is going on.

7.6.3. *The wave function for bonding and anti-bonding states*

With $k = k_p$ and $N = 8$, equation [7.19] for the wave functions can be written as:

$$\psi_{k_p}(x) = c_0 \sum_{t=0}^{8} \exp(ik_p ta)\ \psi_0(x - ta)$$

where $c_0 = 1/\sqrt{N}$ after normalization.

The following example of this function concerns the case where $\psi_0(x - ta) = \psi_t(x)$ correspond to the *s* state (for which the wave functions φ_s correspond to the quantum number $l = 0$, such that the notation for $\psi_0(x - ta) = \psi_t(x)$ is written as $\psi_t(x) = \varphi_{st}(x)$, where $\varphi_{st} = A\mathrm{R}_{n,1=0}(x) = Ce^{-\alpha x}$ (see section 7.1.1).

NOTE.– in equation [7.19], the letter *t* is used to denote the number of atoms in the chain rather than the previously used letter *s*, which is now reserved for use with the φ_s orbital (*s* state characterized by $l = 0$). We can also note that $t = 0 \equiv 8$ as these two values “close” the circle, as shown in Figure 7.22.

7.6.3.1. *Atoms without interactions (where N =8)*

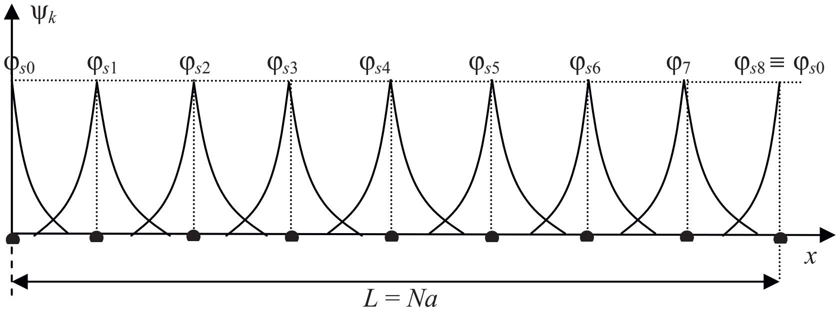


Figure 7.24. *Wave function for a chain of atoms (N = 8) that do not interact and show s states*

Figure 7.24 gives the initial form of the wave function for each *s* state of each non-interacting atom in the chain.

7.6.3.2. *Representation of the ψ_{k_p} functions at the base (bonding state) and at the top (anti-bonding state) of the bond for L = Na = 8a*

States at the bottom of the band where p = 0. When $p = 0$, $k_p = k_0 = 0$ and $\exp(i\, k_p ta) = 1$ whatever value *t* takes on, we thus have (to within the normalization coefficient c_0):

$$\psi_{k_0} = \varphi_{s1} + \varphi_{s2} + \varphi_{s3} + \varphi_{s4} + \varphi_{s5} + \varphi_{s6} + \varphi_{s7} + \varphi_{s8\equiv 0}$$

This is represented in Figure 7.25.

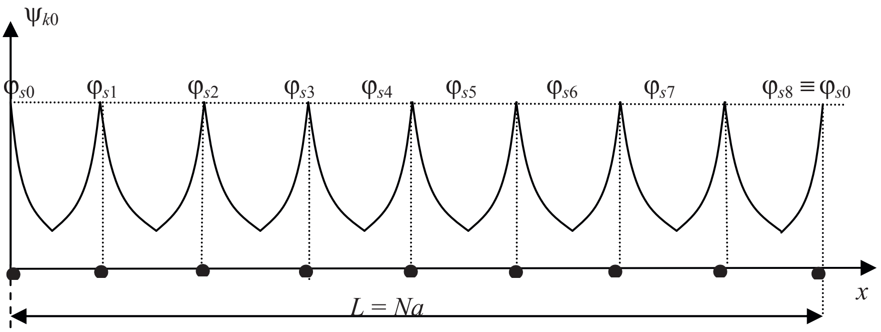


Figure 7.25. *Representation of* ψ_{k_0}

States at the top of the band, where p = 4. Here $k_p = k_4 = \frac{\pi}{a}$. The successive values of *t*, $k_4ta = \pi t$, and $(i\pi t) = \cos(\pi t)$ are given in the table below.

T	0 ≡ 8	1	2	3	4	5	6	7
$k_4ta = \pi t$	0 ≡ 4π	π	2π	3π	4π	5π	6π	7π
cos (π*t*)	1	−1	1	−1	1	−1	1	−1

We also have $\lambda_4 = \frac{2\pi}{k_4} = 2a$, and ψ_{k_4} is represented in Figure 7.26.

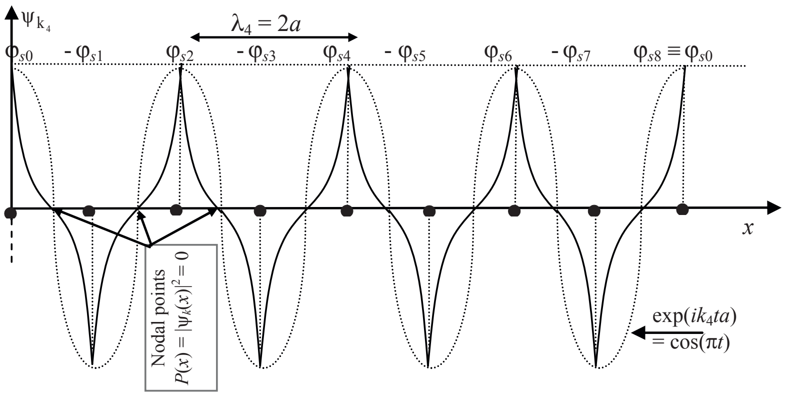


Figure 7.26. *Representation of* ψ_{k_4} *where the atomic wave function* (φ_s) *exhibits a sinusoidal modulation, in this case based on the function cos* πt

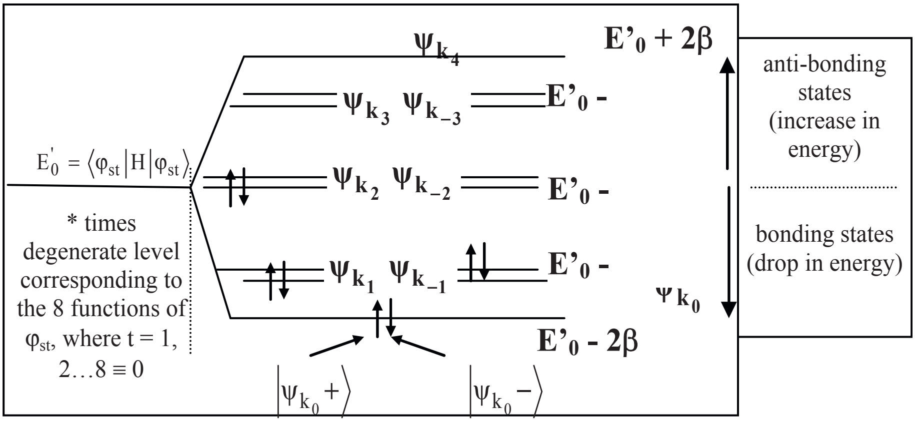


Figure 7.27. *Energy levels and states in a chain where N = 8*

Bonding and anti-bonding states. For the states at the bottom of the band, where $k_0 = 0$, all the coefficients $\exp(ik_0ta)$ are equal (to one) and all the atoms are in phase (Figure 7.25). There are no nodal points in the resulting wave function Ψ_{k0}, which explains why this is the most bonding state. The electron presence probability is reinforced between each atom.

As k_p increases, the energy increases and nodal points start appearing in the wave function. This can be seen in the first figure of problem 1 at the end of this chapter. There, the real component of the wave function ψ_{k1} for k_1 is shown.

In the middle of the band, when $k_2 = \frac{\pi}{2a}$, the states are neither bonding nor anti-bonding and this is the part shown in the second figure of problem 1, which looks at the real part of ψ_{k2}.

In contrast to the above, for the states at the top of the band (where $k_4 = \frac{\pi}{a}$ and $k_4a = \pi$), the successive values of $\exp(i\pi t) = \cos(\pi t)$ alternate between 1 and –1, which results in nodes midway between the atoms (see Figure 7.26). The presence probability for electrons is zero between atoms and they are distributed according to an anti-bonding combination which, evidently, results in an anti-bond.

In terms of energy, and as in Figure 7.27, the bonding states (Ψ_{k0},Ψ_{k1},$\Psi_{k\text{-}1}$) correspond to a drop in energy (a more stable state). In contrast, the anti-bonding states (Ψ_{k4},Ψ_{k3},$\Psi_{k\text{-}3}$) are associated with an increase in energy and are more unstable states.

The energy levels for Ψ_{kp} and $\Psi_{k\text{-}p}$ are identical. As $k_p = -k_{\text{-}p}$, we have $\cos k_p a = \cos[k_{\text{-}p}a]$, and $E_{kp} = E_{k\text{-}p}$ according to equation [7.31]. The two functions are simply associated in terms of the directions of waves' propagation. As for a free electron, the physical solutions can be retained using cos(Re{exp[ik_pta]}), which is represented in the two figures of the aforementioned problem 1.

Most notable is that the interaction between neighboring atoms increases the degeneration as there are eight levels obtained, each associated with a ψ_{k_p} function. Here, the rise in degeneration is only partial, because as indicated above, the two different functions, Ψ_{kp} and $\Psi_{k\text{-}p}$ are at the same energy level.

When incorporating spin, each ψ_{k_p} function gives rise to two functions, namely: $\left|\psi_{k_p}+\right\rangle$ and $\left|\psi_{k_p}-\right\rangle$. Considering that each atom gives one electron to the bond, and that there are eight atoms, hence there are eight electrons to share throughout the energy levels. Thus only the eight states in the lower half of the permitted band are occupied (as in Figure 7.27). The permitted band is half-filled, and therefore, *a priori*, the material is a conductor. There could be a Peierls metal–insulator transition, as detailed in previous sections.

7.6.4. *Generalization to any type of state in an atomic chain*

Section 7.6.3 looked at a system with just *s*-type states. This example can be extended to other states, most notably *p*-type, and the most classic of these representations are treated in the problems at the end of the chapter.

7.7. Conclusion

The energy bands in a periodic lattice are determined by the degree of overlap of the contributing atomic orbitals. The overlap in turn is a product of the competition between the period of the lattice (and the evolution of the potential generated by the atoms) and the value of the radius R of these atomic orbitals.

The semi-free electron model is particularly appropriate when the overlap is sufficiently high for the atoms to lose their own identity (weak bond when a $< 2R$). Chapter 3 is particularly relevant to this.

In the case of strong bonds, we find a more chemical-like representation of the bond between atoms. While the space between atoms is considerable, and the local periodic potential varies sufficiently for atomic orbitals specific to each atom to retain their identity, the bonding orbitals are tied through their linear combination. If a single type of orbital exists, then the energy that is evolved is as a permitted band, with a height equal to 4β, where β is the coupling between adjacent atoms.

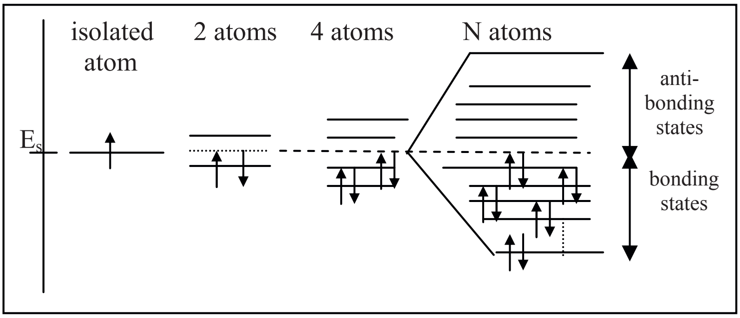


Figure 7.28. *The energy levels of a chain of atoms where each atom contributes an s electron and results in the formations of an energy band without a gap*

If the only atomic orbitals under consideration are *s*-type, the states at the base of the band are bonding combinations of orbitals, whereas the states at the top of the band are anti-bonding orbitals (see Figure 7.28, with an *s*-electron per atom). The amplitude of the wave function is modulated by $\exp(ik_pta)$. When $p = 0$, $k_p = 0$ and the wavelength $\lambda_p = 2\pi/k_p$ tends towards infinity. Therefore there are no nodes if the states is bonding. When $k_p = \pi/a$, then $\lambda_p = 2a$ and, between two adjacent atoms, the atomic orbitals are in opposite phases with nodes between each atom, and thus display what is an essential characteristic of anti-bonds. Problem 2 shows that this characteristic behavior of σ–*s*-bonds appears in π–*p*-bonds.

If there are *N* atoms, then there are *N* levels in all, with *N*/2 bonding. As two electrons can be placed per level (given the effect of spin), the *N* electrons (be they *s*-electrons in σ-orbitals, or *p*-electrons in π-orbitals) will fill *N*/2 of the bonding levels. *N*/2 of the anti-bonding levels will remain empty. It is worth noting that the permitted band will still be equal to 4β, so the higher that *N* is, the closer the levels will be.

Chapter 8 details the extension of this 1D model to the 3D model. The sizes of the permitted bands will be tied to the coordination number (number of bonds for the given atom) of the alternating systems.

Given that the alternation of bond lengths in a 1D chain results in a gap in the middle of the band (as in Figure 7.20), it is also notable that when different types of orbitals interact, then there is also a gap in the energy band. In addition, if during the formation of a solid state, there are two types of orbitals (*s* and *p* for example) that fuse to give hybrids, then once again there is a formation of bonding and anti-bonding orbitals. The difference in energy between bonding and anti-bonding orbitals generates the energy gap. This problem, using the example of sp^3 orbitals, will also be treated in Chapter 8.

7.8. Problems

7.8.1. *Problem 1: complementary study of a chain of s-type atoms where N = 8*

Trace the real part of the wave functions for $k = k_1 = \dfrac{\pi}{4a}$, and $k = k_2 = \frac{\pi}{2a}$.

Answer

This scenario was delved into in section 7.6.3. In the chain are eight atoms, each in an *s*-state. The energy levels are presented in Figure 7.27. The representative functions of the bonding states at the base of the band (ψ_{k0} function) and the anti-bonding states at the top of the band (ψ_{k4} function) are shown in Figures 7.25 and 7.26, respectively. In addition, between $k = k_0$ and $k = k_4$ there are the intermediate states. These have been shown for $k = k_1$ (in Figure 7.29) and $k = k_2$ (in Figure 7.30). ψ'_{k_1} And ψ'_{k_2}, the real parts the wave functions, are such that:

$$\psi'_{k_p}(x) = R\,(\psi_{k_p}(x)) = \mathrm{c}'_0 \sum_{t=0}^{8} \cos\,(k_p ta)\;\psi_0(x-ta)$$

where $c'_0 = \sqrt{\frac{2}{N}}$, $k_p = 2\pi \frac{p}{Na}$ and $\psi_0(x - ta) = C\,e^{-\alpha x}$.

– When $p = 1$, we have $k_1 = \frac{\pi}{4a}$ and $\lambda_1 = \frac{2\pi}{k_1} = 8a$.

The successive values of t, $k_1 ta = \frac{\pi}{4}t$, and $\cos\frac{\pi}{4}t$ are given in the table below.

T	$0 \equiv 8$	1	2	3	4	5	6	7
$k_1 ta = \dfrac{\pi}{4}t$	0	$\dfrac{\pi}{4}$	$\dfrac{\pi}{2}$	$\dfrac{3\pi}{4}$	π	$\dfrac{5\pi}{4}$	$\dfrac{3\pi}{2}$	$\dfrac{7\pi}{4}$
$\cos\dfrac{\pi}{4}t$	1	$\dfrac{\sqrt{2}}{2} = 0.707$	0	–0.707	–1	–0.707	0	0.707

From this can be deduced (to within c'_0) the expression for ψ'_{k1}:

$$\psi'_{k_1} = \varphi_{s0\equiv 8} + 0.707\ \varphi_{s1} + 0\ \varphi_{s2} - 0.707\ \varphi_{s3} - \varphi_{s4} - 0.707\ \varphi_{s5} + 0\ \varphi_{s6} + 0.707\ \varphi_{s7}$$

We thus find that when $k = k_1$, the nodal points start to appear at the level of the wave function shown in Figure 7.29.

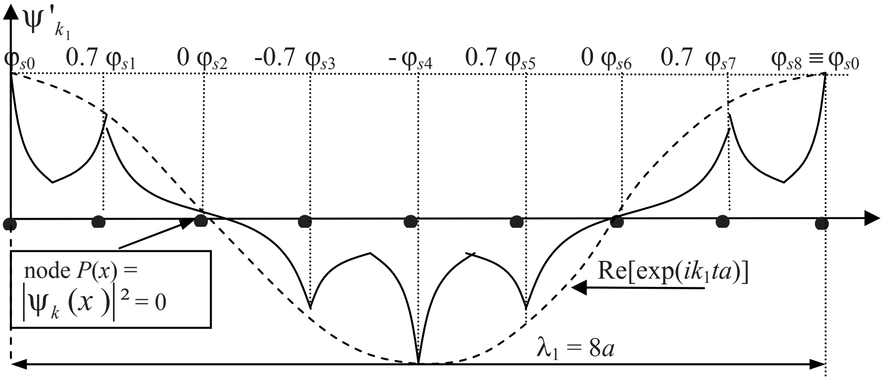


Figure 7.29. *Representation of* ψ'_{k_1} *with:*
$\psi'_{k_1} = R_e(\psi k_1) = \varphi_{s0\equiv 8} + 0.707\varphi_{s1} + 0\varphi_{s2} - 0.707\varphi_{s3} - \varphi_{s4} - 0.707\varphi_{s5} + 0\varphi_{s6} + 0.707\varphi_{s7}$

– When $p = 2$, we have $k_2 = \dfrac{\pi}{2a}$, and $\lambda_2 = \dfrac{2\pi}{\pi/2a} = 4d$

The successive values of t, $k_1\,ta = \frac{\pi}{2}t$, and $\cos\frac{\pi}{2}t$ are given in the table below.

t	$0 \equiv 8$	1	2	3	4	5	6	7
$k_1\,ta = \dfrac{\pi}{2}t$	$0 \equiv 4\pi$	$\dfrac{\pi}{2}$	π	$\dfrac{3\pi}{2}$	2π	$\dfrac{5\pi}{2}$	3π	$\dfrac{7\pi}{2}$
$\cos\dfrac{\pi}{2}t$	1	0	–1	0	1	0	– 1	0

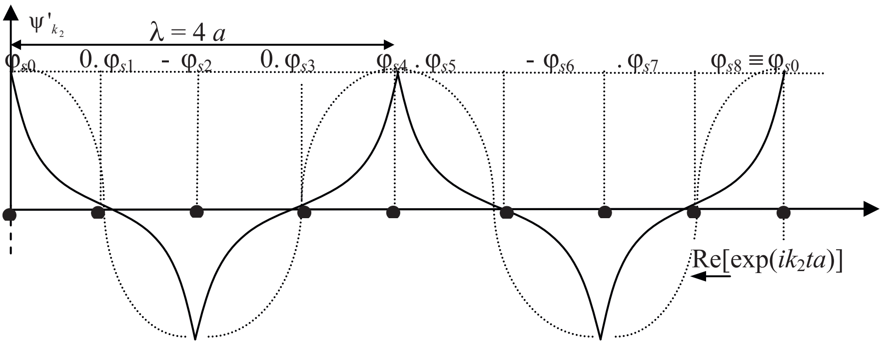


Figure 7.30. *Representation of* $\psi'k_2$ *with*
$\psi'k_2 = R_e(\psi k_2) = \varphi_{s0\equiv8} + 0\varphi_{s1} - \varphi_{s2} + 0\varphi_{s3} + \varphi_{s4} + 0\varphi_{s5} - \varphi_{s6} + 0\varphi_{s7}$

Finally, we can see that in the middle of the band when $k = k_2$, the states are neither bonding nor anti-bonding (Figure 7.30).

7.8.2. *Problem 2: general representation of the states of a chain of σ–s-orbitals (s-orbitals giving σ-overlap) and a chain of σ–p-orbitals*

1)

a) For a chain of atoms from which the *s*-electrons give rise to σ-orbitals (see Figure 7.4), determine the energetic level of bonding and anti-bonding states. Give the band scheme.

b) Show, using the appropriate scheme, how a qualitative result can be found by taking into account the phase associated with the wave function of each atom in the chain.

2)

a) Deal with the same problem but using atoms in a *p*-configuration (see Figure 7.4) that give rise to a σ-type overlap. Add to the figure of question 1a) the scheme for the corresponding band.

b) Same question as 1b) but this time for a chain of σ-*p*-orbitals.

Answers

1)

a) For *s*-orbitals, which are always positive when alone as one *s*-state per atom, then in the atomic chain we have from equation [7.29],

$$-\beta_s = \left\langle \psi_0(x-ta) \right| \mathrm{W}(x-sa) \left| \psi_0(x-sa) \right\rangle = \left\langle \psi_t \right| W \left| \psi_s \right\rangle < 0$$

as ψ_t and ψ_s have the same positive sign (orbital on a single atom *t* or *s*) while $W < 0$. In terms of energy, the result is that from equation [7.31], where *E* is now denoted E_s:

$$E_s = E_{0s} - \alpha_s - 2\,\beta_s \cos ka.$$

From the lower half of Figure 7.31, the energy of the *s*-orbitals:

– drops at the bottom of the band as $-\beta_s$ is always negative and $\cos ka > 0$ ($-\pi/2 \le ka \le \pi/2$ at the bottom of the band) and thus is a bonding state (σ–*s*-band); and

– increases at the summit of the band as $-\beta_s$ is always negative while $\cos ka < 0$ ($\pi/2 \le ka \le \pi$ and $-\pi \le ka \le -\pi/2$ at the top of the band), and thus is an anti-bonding state (σ*–*s*-band).

b) This result can be found by taking the phase term associated with each wave function on each atom into account (in Floquet's development this means the $\cos k.r_s$ term):

– for *s*-type bonding orbitals, the orbitals of two adjacent states are in phase ($k = 0$ as in Figure 7.25), and have the same sign. This situation can also be represented using the sign and geometric shape of the orbital, as in Figure 7.32a. The interaction, or rather, coupling, of the two states decreases the σ-*s*-band energy; and

– for their part, the anti-bonding states are adjacent to orbitals that have an opposite phase (when $k = \pm\, \pi/a$), and hence an opposite sign, as in Figure 7.32b. Their interaction results in an increase in the energy (disfavored state) as shown in Figure 7.31 for the σ*-*s*-band.

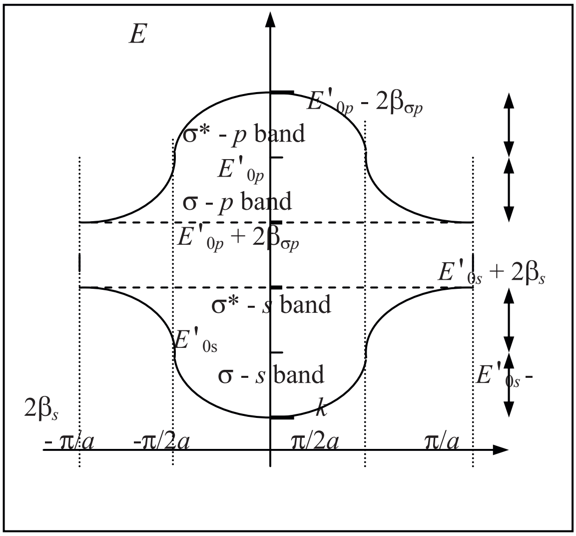


Figure 7.31. *E = f(k) plots for σ-overlaps from s- and p-orbitals*

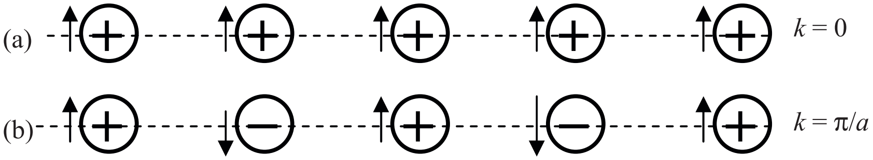


Figure 7.32. *Distribution of (a) bonding σ–s-orbitals; and (b) anti-bonding σ*–s-orbitals; the arrows indicate the direction of the phase*

2)

a) General representation of the states in a chain of σ–p-orbitals (σ-overlap of *p*-orbitals)

Once again, we find for *p* = orbitals that:

$$\beta_{\sigma p} = \left\langle \psi_0(x-ta) \middle| W(x-sa) \middle| \psi_0(x-sa) \right\rangle = \left\langle \psi_t \middle| W \middle| \psi_s \right\rangle \text{ with } W < 0.$$

Now ψ_t and ψ_s represent *p*-orbitals and take on the shape shown in Figure 7.33, that is of a positive lobe adjacent to a negative lobe, so that $-\beta_{\sigma p}$ is positive (upper half of Figure 7.31) and now:

– when $-\pi/2 \leq ka \leq \pi/2$, $\cos ka > 0$ and the energy is increased with respect to $E'_{0p} = E_{0p} - \alpha_p$;

– when $\pi/2 \leq ka \leq \pi$ and $-\pi \leq ka \leq -\pi/2$, we have $\cos ka < 0$ and the energy is decreased with respect to $E'_{0p} = E_{0p} - \alpha_p$.

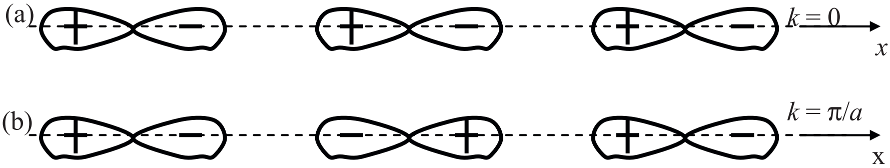


Figure 7.33. *Distribution of (a) anti-bonding σ*–p-orbitals; and (b) bonding σ–p-orbitals*

b) This result can be found by considering that:

– when $k = 0$, the wave functions are in phase and can be shown geometricly as in Figure 7.33a with the example of p_x orbitals. With respect to the Floquet development, the term $\cos k_p sa$ is equal to 1. A positive lobe interacts with a negative lobe forming an anti-bonding state with an increase in energy (σ*–*p*-band shown in Figure 7.31). This behavior is opposite to that of the *s*-band in the same *k* region.

– when $k = \pm\, \pi/a$, the phases alternate ($\cos k_p sa$ alternately equals +1 and –1) so that the geometry can be shown as in Figure 7.33b. The same type of lobes interact and result in the bonding σ–*p*-band shown in Figure 7.31.

7.8.3. *Problem 3: chains containing both σ–s- and σ–p-orbitals*

A chain of atoms contains both σ–*s*- and σ–*p*-bonds. Taking the geometric shape of the *s*- and *p*-orbitals, along with their distributions, show with respect to $k = \pi/a$ the formation of the forbidden band. The top and the bottom of the forbidden band are linked to the probability of electron presence due to the overlap of *s*- and *p*-orbitals.

Answer

As Chapter 3 showed, with respect to $k = \pi/a$, a system of stationary waves can easily arise with wave functions of solutions for the probabilities of presence given by:

– $\psi^+ \propto \cos(\pi x/a)$ and $P^+ = \cos^2(\pi x/a)$. The electronic charges are essentially in the neighborhood of the lattice nodes.

– $\psi^- \propto \sin(\pi x/a)$ and $P^- = \sin^2(\pi x/a)$. The electronic charges are essentially midway between the lattice nodes.

Given the geometric shape of the *s*- and *p*-orbitals and their corresponding electronic distributions, we can state that:

– the probability P^+ concentrates the electronic charges associated with *s*-orbitals around the atomic nodes; and

– the probability P^- concentrates the electronic charges associated with p_x-orbitals at points midway between atoms.

In terms of energy, these two orbitals (and their associated electronic concentrations) are separated by the gap in energy denoted E_G that appears at exactly $k = \pi/a$.

In addition, as the ψ^+ waves associated with the *s*-states concentrate the electrons at the centers of the atoms, the electrostatic interaction energy, at best, decreases the energy of these electrons making them more stable.

For their part, the ψ^- waves associated with *p*-states concentrate their electrons midway between atoms (i.e. at the greatest distance apart) so that the electrons are held furthest from electrostatic interactions with the nuclei of the atoms. This increases their energy making them less stable than the electrons associated with the ψ^+ wave due to *s*-states.

We have thus shown that an energy gap can appear in 1D when two (or more) types of orbital are present. Figure 7.31 can be compared with Figure 7.16, which gives details for σ–*s*- and σ–*p*-orbitals.

7.8.4. *Problem 4: atomic chain with π-type overlapping of p-type orbitals: π–p- and π*–p-orbitals*

For the p_z-type orbitals tied to one another by π-bonds, as shown in Figure 7.34, give a geometric representation of the bonding and anti-bonding orbitals in a chain.

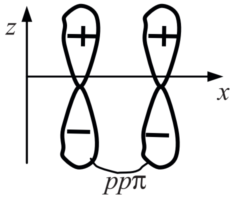


Figure 7.34. *π–p-orbital*

Answer

The resonance integral, $-\beta_{\pi p}$, essentially includes interventions from the overlap of two positive lobes and two negative lobes, as shown in Figure 7.34. When $W < 0$, then $-\beta_{\pi p}$ is negative (just as $-\beta_s$), resulting in a behavior similar to that of σ–*s*-bonds. So Figures 7.35a and b show the geometric representation of bonding (low energy) and anti-bonding (high energy) orbitals.

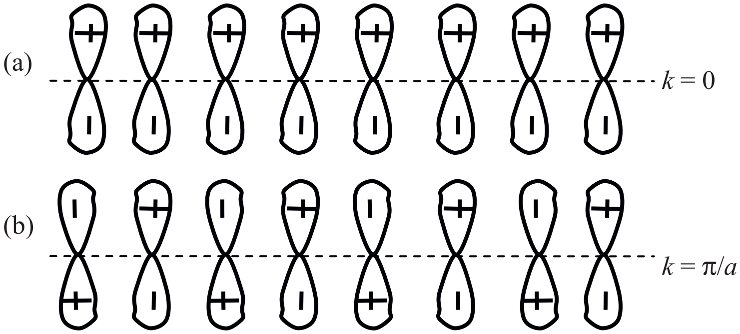


Figure 7.35. *Distribution of (a) bonding π–p-orbitals; and (b) anti-bonding π*–p-orbitals*

Chapter 8

Strong Bonds in Three Dimensions: Band Structure of Diamond and Silicon

This chapter is devoted to the band structure in periodic, 3D solids and contains three studies. The first is on a periodic 3D lattice that has an atom at each node contributing one *s*-electron.

The cubic lattice will be looked at in detail and there is a determination of the height of the permitted (valence) band its effects on the charge mobilities. At this level, the forbidden bands do not yet appear as they require a system with several atomic levels (see Chapter 7 for a comparable study using a 1D system).

The second study focuses on strong covalent bonds in the carbon diamond structure, where each atom is at the center of a tetrahedral and sp^3 hybridization, of which the details of the associated wave functions will be given. This system will then be used to demonstrate the structure of permitted and forbidden bands, as the carbon atoms involve several orbitals.

The third study looks at the band structure of a 3D lattice based on carbon diamond. We will, show in particular, the origin of the band structure with a generation of the forbidden band that separates the bonding and anti-bonding bands (with examples of diamond and silicon).

8.1. Extending the permitted band from 1D to 3D for a lattice of atoms associated with single *s*-orbital nodes (basic cubic system, centered cubic, etc.)

8.1.1. *Permitted energy in 3D: dispersion and equi-energy curves*

8.1.1.1. *General expression*

Equation [7.29] that describes the energy (E) of an electron in a strong bond for a 1D system, can also be written as:

$$E = E_0 - a - b \sum_{t=-1,+1} e^{-ikta}. \qquad [8.1]$$

This summation brings in the two nearest atoms on either side of the atom under study. Most simply, equation [8.1] can be rewritten for more than one dimension as:

$$E = E_0 - a - b \sum_{m} e^{-i\vec{k}\,\overrightarrow{a_m}} \qquad [8.2]$$

where $\overrightarrow{a_m}$ represents the vectors joining the reference atom to its nearest "*m*" neighbors.

8.1.1.2. *Expression for energy in the simple cubic system: equi-energy and dispersion curves*

For a simple cubic lattice, as in Figure 8.1a, the $\overrightarrow{a_m}$ vectors obtained for nearest neighbors, have the components:

($\pm a$, 0, 0) in x
(0, $\pm a$, 0) in y
(0, 0, $\pm a$) in z

The energy thus takes on the form:

$$E = E_0 - \alpha - 2\beta\,[\cos k_x a + \cos k_y a + \cos k_z a] \qquad [8.3]$$

where k_x, k_y, k_z are the components for $\vec{k}$ in the three direction Ox, Oy, Oz.

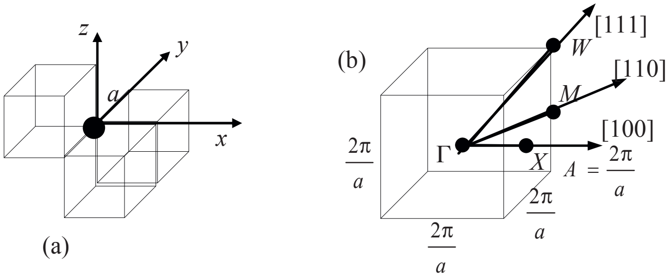


Figure 8.1. *(a) Structure of the cubic direct lattice; (b) Brillouin zone in the reciprocal lattice*

8.1.1.2.1. Dispersion curves for the simple cubic system

– At the center of the zone, i.e. the point Γ where $k = 0$ (and $k_x = k_y = k_z = 0$), the energy is at a minimum and is given by:

$$E = E_0 - \alpha - 6\beta = E(\Gamma) \qquad [8.4]$$

– With respect to direction [100], we have $k_x = k$, $k_y = 0$ and $k_z = 0$, from which, according to equation [8.3], $E_{[100]} = E_0 - \alpha - 4\beta - 2\beta\cos k_x a$, and thus in X, where $\overrightarrow{\Gamma X}\begin{cases} k_x = \pi/a \\ k_y = 0 \\ k_z = 0 \end{cases}$, we have $\cos k_x a = -1$ and hence $E(X) = E_0 - \alpha - 2\beta$.

– With respect to [110], we have $k_x = k_y = \frac{k}{\sqrt{2}}$, $k_z = 0$, so according to equation [8.3], $E_{[110]} = E_0 - \alpha - 2\beta - 4\beta\cos k_x a$, and in M where $\overrightarrow{\Gamma M}\begin{cases} k_x = \pi/a \\ k_y = \pi/a \\ k_z = 0 \end{cases}$, we have, with $\cos k_x a = -1$, the result $E(M) = E_0 - \alpha + 2\beta$.

– With respect to [111], we have $k_x = k_y = k_z = \frac{k}{\sqrt{3}}$ (as $k^2 = 3k_x^2$), so that $E_{[111]} = E_0 - \alpha - 6\beta\cos k_x a$, and in W, where $\overrightarrow{\Gamma W}\begin{cases} k_x = \pi/a \\ k_y = \pi/a \\ k_z = \pi/a \end{cases}$, with $\cos k_x a = -1$ we have:

$$E(W) = E_0 - \alpha + 6\beta. \qquad [8.5]$$

It should be noted that in $W, k = k_x\sqrt{3} = \frac{\pi}{a}\sqrt{3}$.

The $E(k)$ dispersion curve with respect to the [100] and [111] directions is given in Figure 8.2. It is notable that the size of the permitted band is given by:

$$\Delta E = E(W) - E(\Gamma) = 12\beta.$$

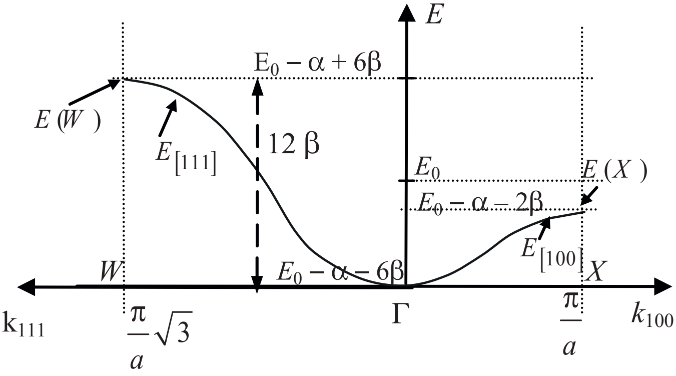


Figure 8.2. *E(k) dispersion curve for the cubic system*

8.1.1.2.2. Equi-energy curves for the simple cubic system

In addition to the calculation shown below, a related and detailed calculation of the equi-energy curves for a 2D simple square lattice can be found in problem 1 at the end of this chapter.

– In the neighborhood of the zone center (Γ), $k \approx k_0 \approx 0$, so that $k_x \approx k_y \approx k_z \approx 0$, and $\cos k_x a \simeq 1 - \frac{(k_x a)^2}{2}$ (and likewise for $\cos k_y$ and $\cos k_z$). So the energy can be written as:

$$E(k \approx k_0) = E_0 - \alpha - 6\beta + \beta a^2 (k_x^2 + k_y^2 + k_z^2)$$
$$= E(\Gamma) + \beta k^2 a^2. \qquad [8.6]$$

From this we deduce that $\frac{E-E_0+\alpha+6\beta}{\beta a^2} = k_x^2 + k_y^2 + k_z^2 = k^2$. In the reciprocal k space, for a constant E, we have $k = \sqrt{\left(E - E_0 + \alpha + 6\beta\right)/\beta a^2} =$ constant. Thus, in the neighborhood of Γ, the equi-energy surfaces are spheres with radii equal to k as given in the proceeding equation.

– In the plane of $k_z = 0$, equation [8.3] becomes $E = E_0 - \alpha - 2\beta - 2\beta(\cos k_x a + \cos k_y a)$. This means that in this plane, the equi-energy curve $(E^{(1)} = E_0 - \alpha - 2\beta)$ is obtained with cos k_xa + cos k_ya = 0. This condition results in $k_y = \pm k_x \pm \frac{\pi}{a}$ (see problem 1 (section 8.5.1), question 3c), of which the $k_y = f(k_x)$ representation gives lines that generate a square in the plane along $k_z = 0$ and with sides equal to $\frac{\pi}{a}\sqrt{2}$. The summits are at $X(\frac{\pi}{a},0,0)$, $X_1(0,\frac{\pi}{a},0)$, $X_2(-\frac{\pi}{a},0,0)$, $X_3(0,-\frac{\pi}{a},0)$.

By analogy and in the planes defined by $k_x = 0$ and $k_y = 0$, we find the exact equi-energy surface; $E^{(1)} = E_0 - \alpha - 2\beta$ being that of a cube, the six corners of which have the coordinates $(\pm\frac{\pi}{a},0,0)$, $(0,\pm\frac{\pi}{a},0)$, $(0,0,\pm\frac{\pi}{a})$.

– When $E = E_0 - \alpha$ (equi-energy lines $E^{(2)} = E_0 - \alpha$) we should have, according to equation [8.3], cos k_xa + cos k_ya = –cos k_za.

In the plane with side $k_z = \pm\frac{\pi}{2a}$, where cos k_xa + cos k_ya = 0, we should again have $k_y = \pm k_x \pm \frac{\pi}{a}$, of which the $k_y = f(k_x)$ representation in the plane of side $k_z = \pm\frac{\pi}{2a}$ gives straight lines that generate a square (see Figure 8.3).

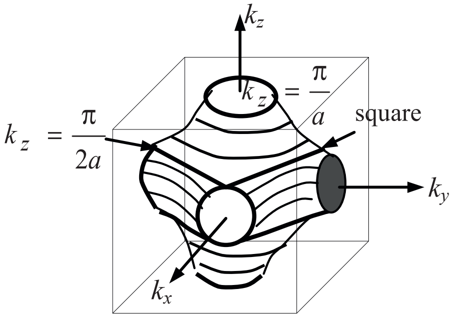


Figure 8.3. *Constant energy surface when $E^{(2)} = E_0 - \alpha$*

In the planes with sides $k_z = \pm\frac{\pi}{a}$, we should have cos k_xa + cos k_ya = 1. Notably, Sommerfeld and Bethe showed (see Kittel in Figure 6, Chapter 11, *Quantum Theory of Solids*, Wiley, 1963) that it is in this place that the equi-energy surface is practically circular. This means that the radius (k) of this surface is constant in all directions of the plane (from which $k_x \approx k_y$), so that $k = \sqrt{k_x^2 + k_y^2} \overset{k_x \approx k_y}{\approx} k_x\sqrt{2}$. As $k_x \approx k_y$ then cos $k_xa \approx$ cos k_ya, so that cos k_xa + cos k_ya = 1 and cos $k_xa \approx$ cos $k_ya \approx \frac{1}{2}$ (and $k_xa \approx k_ya \approx \frac{\pi}{3}$), so that $k \approx k_x\sqrt{2} \approx \frac{\pi\sqrt{2}}{3a}$. This is the radius of the circle associated with the equi-energy line $(E^{(2)} = E_0 - \alpha)$ in the plane of side given by $k_z = \pm\frac{\pi}{a}$.

In the intermediate planes with k_z sides and such that $\frac{\pi}{2a} < k_z < \frac{\pi}{a}$, or $-\frac{\pi}{a} < k_z < -\frac{\pi}{2a}$, the equi-energy lines which start in the shape of a square (in $k_z = \pm\frac{\pi}{2a}$) evolve towards a *pseudo*-circular shape, as represented in Figure 8.3.

8.1.1.3. *Expressions for energy in centered cc and fcc systems – see problem 2*

These calculations, which are rather long, are detailed in problem 2 at the end of this chapter: question 1a for the cubic centered lattice, and 2a for the cubic face centered lattice.

8.1.2. *Expression for the band width*

8.1.2.1. *Reminder of the expression for energy in the neighborhood of k_0 at the center of the zone, obtained through a Maclaurin development of E*

Briefly recalled, the MacLaurin development for the $E(k)$ energy can be written for the neighborhood of k_0 as

$$E_{(k)} = E_{(k_0)} + (k - k_0)\left(\frac{\partial E}{\partial k}\right)_{k_0} + \frac{(k - k_0)^2}{2}\left(\frac{\partial^2 E}{\partial k^2}\right)_{k_0}$$

Remembering that at the center of the zone there is a horizontal tangent, $\left(\frac{\partial E}{\partial k}\right)_{k_0} = 0$, and with the introduction of an expression for the effective mass (see Chapter 4), when $k \approx k_0$ we have for the energy:

$$E_{(k)} = E_{(k_0)} + \frac{\hbar^2}{2m^*}(k - k_0)^2. \qquad [8.7]$$

In the neighborhood of the zone center, where $k_0 \approx 0$, we thus have

$$E_{(k)} = E_{(k_0)} + \frac{\hbar^2}{2m^*}k^2. \qquad [8.7']$$

8.1.2.2. *The simple cubic lattice*

The identification of the coefficients in k^n ($n = 0$ and $n = 2$) for equations [8.6] and [8.7’] yields:

$$\begin{cases} E_{(k_0)} = E_0 - \alpha - 6\beta \text{ (which returns equation [8.4])} \\ \beta a^2 = \dfrac{\hbar^2}{2m^*}, \text{ so that } m^* = \dfrac{\hbar^2}{2\beta a^2}. \end{cases} \qquad [8.8]$$

Incidentally, equation [8.8] shows that the greater the transfer integral (β), that is the easier it is for electrons to pass from one site to an adjacent site, the smaller the effective mass.

The size of the band can be obtained from the amplitude in the variation of the dispersion energy curve (Figure 8.2) which is such that:

– when $k_x = k_y = k_z = 0$, then:

$$E = E\,(k_0 = 0) = E_0 - \alpha - 6\beta \qquad [8.4]$$

– when $k_x = k_y = k_z = \frac{\pi}{a}$ then:

$$E = E\left(k_x = \frac{\pi}{a}, k_y = \frac{\pi}{a}, k_z = \frac{\pi}{a}\right) = E_0 - \alpha + 6\beta \qquad [8.5]$$

From this it can be deduced that:

$$\Delta E = E\left(k_x = \frac{\pi}{a}, k_y = \frac{\pi}{a}, k_z = \frac{\pi}{a}\right) - E(k_0) = 12\beta,$$

so that on using the notation given as $\Delta E = B$, we find:

$$B = 12\beta \qquad [8.9]$$

8.1.2.3. *Generalization*

Looking back on the structures that have been studied up to now, which include the linear chain in Chapter 7, the square place lattice in problem 1, the simple cubic lattice in section 8.1.1.2, and the centered and centered face cubic lattices in problem 2, the most bonding state came from the *s*-orbitals for $k = 0$ (Γ point). The corresponding energy is given by $E\,(\Gamma) = E_0 - \alpha - Z\beta$, where Z is the coordination number that is equal to the number of "closest neighbor" sites. This can number 2 for a linear chain, 4 for a square plane lattice, 6 for a cubic lattice, 8 for a centered cubic (cc) lattice, and 12 for a face centered cubic (fcc) lattice.

For alternating structures, that is all of the aforementioned except for the fcc structure (see comment at the end of section 8.5.2), in which all closed circuits link the close neighbor atoms to a paired number of branches, the most anti-bonding state has an energy given by $E\,(X\,) = E_0 - \alpha + Z\beta$.

For numerous alternating structures, equation [8.9] can therefore be written as

$$B = 2Z\beta \qquad [8.10]$$

Finally, in all cases, be they alternating structures or not, the size of the band is proportional to β. From equation [8.10], for alternating structures, the weaker the β coupling between neighboring atoms, the narrower the permitted bands. Non-alternating structures are dealt with in section 8.5.2.

8.1.3. *Expressions for the effective mass and mobility*

For the effective mass, we note that equation [8.8], $m^* = \frac{\hbar^2}{2\beta a^2}$, established around the Γ point (the origin of the reciprocal lattice) for a simple cubic lattice is also valid for square plane lattices (problem 1, question 4), cc and fcc lattices (problem 2, question 1c and 2c, respectively).

The mobility of electrons with a charge denoted -*q*, and a velocity $\vec{v}$, subject to an external electric field $\vec{E}$, is defined by the relation $\langle \vec{v} \rangle = -\mu\vec{E}$. The integration of the fundamental dynamic equation, over an average time $\langle t \rangle = \tau$, where τ is the relaxation time for electrons colliding with the lattice, written in the form $\vec{F}_{\text{exterior}} = -q\vec{E} = m^* \frac{d\vec{v}}{dt}$, gives $\langle \vec{v} \rangle = -\frac{q}{m^*}\tau\vec{E}$, from which by identification with the defining $\langle \vec{v} \rangle = -\mu\vec{E}$ we have $\mu = \frac{q\tau}{m^*}$.

For alternating structures, the introduction of β deduced from equation [8.10] (β = *B*/2*Z*) and from equation [8.8] gives:

$$m^* = \frac{\hbar^2}{Ba^2}Z. \qquad [8.11']$$

Placing this into *μ*, gives:

$$\mu = \frac{q\tau a^2}{\hbar^2}\frac{B}{Z}. \qquad [8.11]$$

For non-alternating structures, there are similar laws for the variation of m^* and μ as a function of B (only the coefficients are different, see the problem in section 8.5.2), so that $m^* \propto \frac{1}{B}$ and $\mu \propto B$.

We can conclude that the semiconductors with narrow permitted bands i.e. $\Delta E = B$ and β are small due to poor coupling between atoms (equation [8.10]), and are semiconductors with high effective masses and low mobilities.

8.2. Structure of diamond: covalent bonds and their hybridization

8.2.1. *The structure of diamond*

A study of the energy bands that can appear in materials such as carbon diamond requires a determination of the involved orbitals and wave functions.

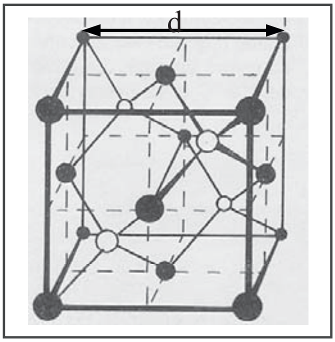


Figure 8.4. *Positions of atoms with the first cell shown in black, the second shown in white*

Diamond has the same crystalline system as silicon (Si) or germanium (Ge). As described in Chapter 6, this system can be reduced to a fcc system as it is based on one fcc imbricated within another. The second cell has its edges parallel to those of the first but is shifted along a diagonal of the cube. This shift is equal to one quarter of a full diagonal (see Figure 6.6b).

Figure 8.4 shows the position of all the atoms in the first cell, those in the first cell coloured black, and those in the second coloured white. This figure shows that each atom has four immediate and symmetrically placed neighbors. It is in effect placed in the center of a regular tetrahedral, the top of which is occupied by a neighboring atom. The cell has the dimensions of d = 3.56 Å for diamond, d = 5.43 Å for silicon, and d = 5.62 Å for germanium. The distance between two neighboring atoms is equal to $d\big/\sqrt{3/16}$, i.e. 3.08 Å for C, 4.7 Å for Si and 4.86 Å for Ge.

As the bonds between the reference atom and its four neighbors are equally placed in space, they are all the same and necessitate the hybridization of the original atomic orbitals. Before going into this so-called sp^3 hybridization in detail (in section 8.2.3), the next section deals with the fundamentals of hybridization.

8.2.2. *Hybridization of atomic orbitals*

8.2.2.1. *The notation of hybridization*

The carbon atom has the atomic configuration $1s^2\ 2s^2\ 2p^2$. It should act as a divalent element, such as in CO_2, but this configuration does not explain the formation of CH_4 (sp^3 tetragonal hybridization), of $CH_2{=}CH_2$ (sp^2 triagonal hybridization), or of C_2H_2 (sp^1 diagonal hybridization).

This behavior can be interpreted using the example of two carbon atoms "uniting". Each atom has valence states, 2*s* and 2*p*, that are represented by one *s*-orbital and three *p*-orbitals. This first excited state, $1s^2\ 2s^1\ 2p^3$, is favored if the resulting energy levels show a drop, as is the case when these orbitals "mix", or rather, are hybridized.

Such molecular orbitals can be denoted in the following way, where each carbon atom *A* and *B* is identified using the numbers 1 and 2, respectively:

$$
\begin{aligned}
|\Psi> = {} & A_1\,|2s>_1 + A_2\,|2p_x>_1 + A_3\,|2p_y>_1 + A_4\,|2p_z>_1 \\
& + B_1\,|2s>_2 + B_2\,|2p_x>_2 + B_3\,|2p_y>_2 + B_4\,|2p_z>_2 \\
= {} & |\Psi_1> + |\Psi_2>.
\end{aligned}
$$

8.2.2.2. *Various levels of hybridization*

There are in fact different levels of coupling that can be attained, each bringing into play stronger or weaker couplings, or in other terms, varying levels of hybridization.

8.2.2.2.1. The sp^1 diagonal hybridization

This is due to coupling between s and p_x, for example:

$$|\Psi > = A_1|2s >_1 + A_2|2p_x >_1 + B_1|2s >_2 + B_2|2p_x >_2$$

$$= |\Psi_1 > + |\Psi_2 >$$

where $|\Psi_1 >$ and $|\Psi_2 >$ each correspond to two possible orbitals. In all this gives two bonding orbitals (σ_ℓ) and two anti-bonding orbitals $(\sigma_a{}^*)$. Therefore, $|\Psi_1 >$ can be such that $\left|\psi_{1\ell}\right\rangle = a_1|2s >_1 \pm a_2|2p_x >_1$ and $|\Psi_{1a} > = a'_1|2s >_1 \pm a'_2|2p_x >_1$, with the + and – sign being attributed as a function of the sign of the hybridizing orbitals.

The remaining π-orbitals that have not been used during the sp^1 hybridization process are thus obtained only in the directions (Oy) and (Oz), as shown in Figure 8.5.

8.2.2.2.2. The triagonal sp^2 hybridization

This hybridization leaves, for example, the $|2p_z>$ state outside of a linear combination which only involves the $|2s >$, $|2p_x >$ and $|2p_y>$ states. In effect,

$$\Psi = A_1|2s >_1 + A_2|2p_x >_1 + A_3|2p_y >_1 + B_1|2s >_2 + B_2|2p_x >_2 + B_3|2p_y >_2$$

$$= |\Psi_1 > + |\Psi_2 >$$

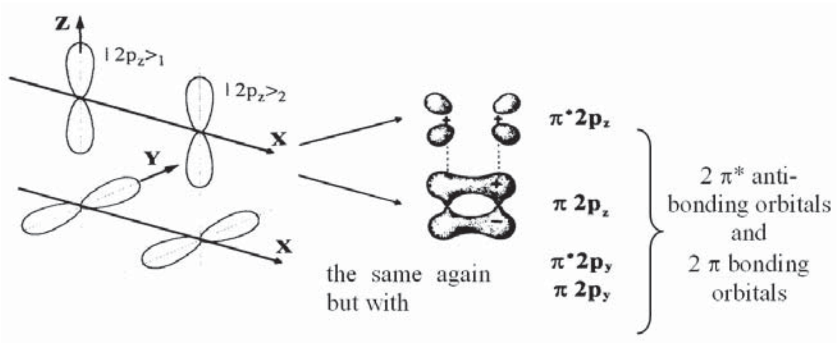


Figure 8.5. *π*- and π-orbitals*

The three hybrid orbitals have their axes in the same plan and are generally denoted $2sp_a^2$, $2sp_b^2$, $2sp_c^2$. The fourth orbital, $2p_z$, remains as it is in a plane perpendicular to that above, and gives rise to a π-orbital.

So, for this molecule:

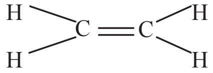


Figure 8.6 gives a schematic illustration.

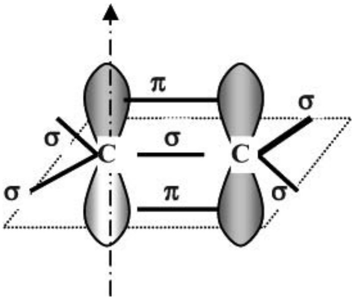


Figure 8.6. *Ethene (C_2H_4) orbitals*

The angle between the three hybrid σ-bonds on the same carbon atom must equal 120. A calculation that is identical to that developed in section 8.2.3 (for sp^3 hybridization) makes possible a determination of the three sp^2 orbitals:

$$h_1 = \frac{1}{3^{1/2}}\left(\left|s\right\rangle + 2^{1/2}\left|p_x\right\rangle\right), \; h_2 = \frac{1}{3^{1/2}}\left(\left|s\right\rangle - \frac{\left|p_x\right\rangle}{2^{1/2}} + \frac{3^{1/2}}{2^{1/2}}\left|p_y\right\rangle\right),$$

$$h_3 = \frac{1}{3^{1/2}}\left(\left|s\right\rangle - \frac{\left|p_x\right\rangle}{2^{1/2}} - \frac{3^{1/2}}{2^{1/2}}\left|p_y\right\rangle\right).$$

8.2.2.2.3. The tetragonal sp^3 hybridization

This hybrid brings in all states (at least of those we have looked at) without exclusion. As noted above it is this hybridization that is present in diamond. Given its importance and prevalence for elements in the IV column of the periodic table, it will be dealt with in more detail in the following section.

8.2.3. *sp*^{3} *Hybridization*

8.2.3.1. *Orbital revolution symmetry*

The normalized wave functions of the *s* and *p* states (where $r \approx r_f$ for the expansion zone for the orbitals) of the valence electrons of carbon (n = 2) are the result of preliminary calculations already performed in Chapter 7 (see section 7.1.1, where the atomic wave functions are denoted ψ, while here they are denoted using φ and ψ is now reserved for hybrid orbitals), as in:

$$\langle s| = \mathrm{R}_{\mathrm{n},0}\,(\mathrm{r})\,\Theta_{0,0}\,\Phi_0 = h_{(r_f)} = S\,(n = 2, \mathrm{l} = 0, m = 0)$$

$$|p_z\rangle = \varphi_0 = \frac{1}{2}\sqrt{\frac{3}{\pi}}R\,(r_f\,)\cos\theta = g\,(r)\cos\theta = zf\,(r) = \mathrm{Z} = \varphi_{2p_z}$$
$$(n = 2, \mathrm{l} = 1, m = 0)$$

$$|p_x\rangle = \varphi_1 = \frac{1}{2}\sqrt{\frac{3}{\pi}}R\,(r_f\,)\sin\theta\cos\varphi = g\,(r)\sin\theta\cos\varphi$$
$$= x\,f\,(r) = \mathrm{X} = \varphi_{2p_x} \quad (n = 2, \mathrm{l} = 1, m = 1)$$

$$|p_y\rangle = \varphi_{-1} = \frac{1}{2}\sqrt{\frac{3}{\pi}}R\,(r_f\,)\sin\theta\sin\varphi = g\,(r)\sin\theta\sin\varphi$$
$$= y\,f\,(r) = Y = \varphi_{2p_y} \quad (n = 2,\ l = 1,\ m = -1)$$

The $f(r)$ and $g(r)$ functions are related by the simple equation $g(r) = r\,f(r)$.

The four proper functions, S, X, Y and Z, make up the orthonormalized base in a space with four dimensions.

For the carbon sp^3 hybrid state, in which the S, X, Y, and Z orbitals play no particular role, a new base can be obtained using the preceding base. It comprises four functions that account for the dimensionality of the system:

$$\begin{cases} \Psi_1 = \alpha_1\, S + a_1\, X + b_1\, Y + c_1\, Z \\ \Psi_2 = \alpha_2\, S + a_2\, X + b_2\, Y + c_2\, Z \\ \Psi_3 = \alpha_3\, S + a_3\, X + b_3\, Y + c_3\, Z \\ \Psi_4 = \alpha_4\, S + a_4\, X + b_4\, Y + c_4\, Z \end{cases}$$

$|\,\Psi_1 >$, $|\,\Psi_2 >$,$|\,\Psi_3 >$,$|\,\Psi_4 >$ are the hybrid orbitals corresponding to the $1s^2\,2s^2\,2p^2 \to 1s^2\,2s^2\,2p^2 \overset{\text{then}}{\to} 1s^2\,2t^4$ that are only possible if the energy states undergo a gain in energy required by the initial excitation of the atomic orbitals.

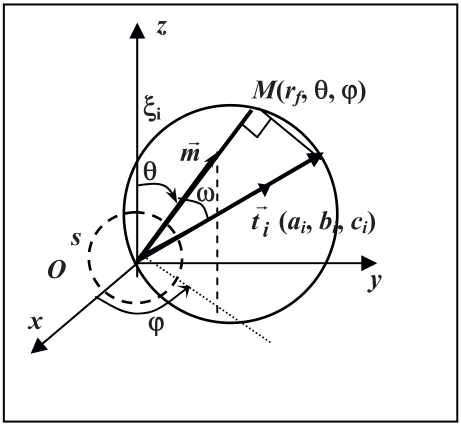


Figure 8.7. *Representation of Ψ_i*

Figure 8.7 illustrates the denoted terms. The unit vector in the direction OM is denoted $\vec{m}$ and the point M has spherical coordinates r_f, θ and φ. In terms of Cartesian coordinates, where $\vec{e}_x$, $\vec{e}_y$, and $\vec{e}_z$ are the unit vectors for the axes Ox, Oy and Oz, the respective components of $\overrightarrow{OM}$ are:

$x = r_f\,\sin\theta\cos\varphi$, $y = r_f \sin\theta\sin\varphi$ and

$z = r_f\,\cos\theta$.

The components of $\vec{m}$ (unit vector where $|r_f| = 1$) in the Ox, Oy, and Oz axes system are thus:

$$\begin{cases} \sin\theta\cos\varphi = \vec{m}.\vec{e}_x \\ \vec{m} \ \sin\theta\sin\varphi = \vec{m}.\vec{e}_y \\ \cos\theta = \vec{m}.\vec{e}_z \end{cases}$$

Taking into account the form of X, Y and Z given at the beginning of this section, it is possible to write that

$$\begin{cases} X = g(r).\vec{e}_x\,.\vec{m} \\ Y = g(r).\vec{e}_y\,.\vec{m} \\ \mathrm{Z} = g(r).\vec{e}_z\,.\vec{m} \end{cases}$$

With $i = 1, 2, 3$, and 4, it is now possible to state that:

$$\Psi_i = \alpha_i\,\mathrm{S} + g(r)\,(a_i\,\vec{e}_x + b_i\,\vec{e}_y + c_i\,\vec{e}_z\,)\,\vec{m}$$

where the vector $\vec{t_i} = a_i\,\vec{e}_x + b_i\,\vec{e}_y + c_i\,\vec{e}_z$ is introduced. It has the components a_i, b_i, and c_i, for the Ox, Oy and Oz axes.

Each Ψ_i function can thus be written as: $\Psi_i = \alpha_i\,S + g(r)\,\vec{t_i}\,.\vec{m}$. With $|\vec{m}| = 1$, $|\vec{t_i}| = \sqrt{a_i^2 + b_i^2 + c_i^2}$ and by making $\omega = (\vec{t_i}\,, \vec{m}\,)$, we find that:

$$\Psi_i = \alpha_i\,S + g(r)\sqrt{a_i^2 + b_i^2 + c_i^2}\,\cos\omega$$

This finally gives $\Psi_i = \alpha_i\,S + \xi_i$, where the ξ_i orbital, thus introduced, is in the form:

$$\xi_i = a_i X + b_i Y + c_i Z = g(r)\,\vec{t_i}\,.\vec{m}$$

Here the ζ_i orbital is collinearly normalized with ξ_i, and therefore is such that $<\zeta_i|\zeta_i> = 1$ and $\xi_i = \lambda_i\, \zeta_i$. ζ_i is thus written as:

$$\zeta_i = \frac{\xi_i}{\lambda_i} = \frac{1}{\lambda_i}(a_i X + b_i Y + c_i Z) = (a'_i X + b'_i Y + c'_i Z)$$

$$= \frac{1}{\lambda_i} g(r)\vec{t}_i .\vec{m} = g(r)\,\frac{\vec{t}_i}{\lambda_i}\vec{m} = g(r)\ \vec{t}_i\,'.\vec{m}\ \text{, with } \vec{t}_i\,' = \frac{\vec{t}_i}{\lambda_i}\begin{cases}\dfrac{a_i}{\lambda_i} = a_i' \\[2ex] \dfrac{b_i}{\lambda_i} = b_i' \\[2ex] \dfrac{c_i}{\lambda_i} = c_i'\end{cases}$$

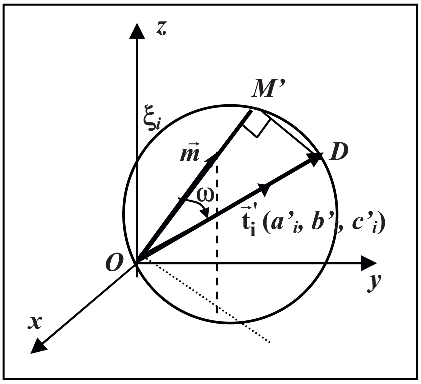


Figure 8.8. *Representation of* ζ_i

As $< \zeta_i \mid \zeta_i > = 1 = a'^2_i + b'^2_i + c'^2_\mathrm{i} = \mid \vec{t}_i\,' \mid^2$ we can state that $\zeta_i = \mathrm{g(r)}\ \vec{t}_i\,'.\vec{m} = \mathrm{g(r)} \cos(\vec{t}_i\,', \vec{m}) = \overrightarrow{OD}.\,\vec{m}$, with $\overrightarrow{OD} = g(r)\ \vec{t}_i\,'$.

With $\omega = (\vec{t}_i\,', \vec{m})$, we have $\zeta_i = OD \cos \omega$, and the M' points are placed such that $\zeta_i = OM'$ is a sphere with a diameter, given by $|\overrightarrow{OD}| = g(r)$, is directed along the unit vector $\vec{t}_i\,'$. Figure 8.8 shows how, in the plane of this sheet of paper, this system is similar to Figure 7.2 that was used to describe p_z-orbitals. Similarly, the M points are placed such that $OM = \xi_i = \lambda_i\, \zeta_i$ is a sphere of radius $\lambda_i\, OD$.

So, using Figure 8.7, the hybrid orbital Ψ_i is obtained by addition of the S orbital (that has a spherical symmetry around O, as shown in the figure) with the orbital denoted by $\xi_i = \lambda_i \zeta_i$, which also revolves around $\vec{t_i}\,'$ or $\vec{t_i}$ $(= \lambda_i \vec{t_i}\,')$. If $\alpha_i > 0$, this addition is wholly positive in magnitude with respect to ξ_i when $\xi_i > 0$.

8.2.3.2. *Angle between the direction of equivalent hybrid orbitals*

This angle can be determined between the two directions around which revolve ξ_i- or ζ_i-type orbitals, with the condition for hybridization being that the Ψ_1 and Ψ_2 orbitals are equivalent. If the two orbitals are denoted 1 and 2, then for example:

$$\Psi_1 = \alpha_1 S + \xi_1 = \alpha_1 S + \lambda_1 \zeta_1 = \alpha_1 (S + \mu_1 \zeta_1) \qquad \text{where } \lambda_1 = \alpha_1 \mu_1$$

$$\Psi_2 = \alpha_2 S + \xi_2 = \alpha_2 S + \lambda_2 \zeta_2 = \alpha_2 (S + \mu_2 \zeta_2) \qquad \text{where } \lambda_2 = \alpha_2 \mu_2$$

The condition of orthogonality between Ψ_1 and Ψ_2 (to an orthonormalized base) gives:

$$\begin{aligned} <\Psi_1 \mid \Psi_2> &= \alpha_1\alpha_2 <S + \mu_1\zeta_1|S + \mu_2 \zeta_2> \\ &= \alpha_1\alpha_2 \,[1 + \mu_1\mu_2<\zeta_1 \mid \zeta_2>] = 0 \end{aligned}$$

where the S and ζ_i functions are orthonormal to each other.

With α_1 and α_2 not being equal to zero, we can deduce that we should see $1 + \mu_1\mu_2 \cos \alpha_{12} = 0$, where $\alpha_{12} = (\vec{t_1}\,', \vec{t}\,'_2)$ represents the angle between the two axes of rotation of the normalized wave functions denoted ζ_1 and ζ_2. The equivalence imposed on the Ψ_1 and Ψ_2 also means that $\mu_1 = \mu_2$. In the case of Ψ_1 and Ψ_2, the S orbitals are similarly equivalent because when using the same coefficient (1) in the brackets, we should find that $\mu_1 = \mu_2 = \mu$ such that $\Psi_1 \equiv \Psi_2$).

Finally, the Ψ_1 and Ψ_2 orbitals are equivalent if $\cos \alpha_{12} = -\dfrac{1}{\mu^2}$.

8.2.3.3. *Wave functions in the sp^3 hybridization*

In this case the carbon is at the center of a tetrahedron and its orbitals are aligned along axes that run between that center and the extremities of the tetrahedron, as shown in Figure 8.9. These lines form angles that can be calculated using, for example, the components $\vec{t_1}\,'$ and $\vec{t}\,'_2$.

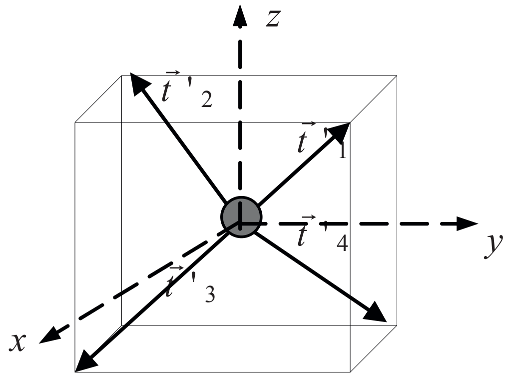


Figure 8.9. *sp^3 hybridization*

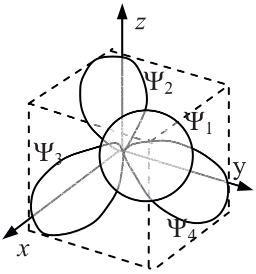
As $|\vec{t}\,'_1| = 1$, and the carbon is at the center of a tetrahedron, we should find that $a'^{2}_1 + b'^{2}_1 + c'^{2}_1 = 1$ and $a'_1 = b'_1 = c'_1$, so that $a'_1 = b'_1 = c'_1 = \frac{1}{\sqrt{3}}$.

Similarly, for $\vec{t}\,'_2$ on taking its geometric position into account, we find that $a'_2 = b'_2 = -\frac{1}{\sqrt{3}}$ and $c'_2 = \frac{1}{\sqrt{3}}$. Thus, $\cos \alpha_{12} = \vec{t}\,'_1 . \vec{t}\,'_2 = -\frac{1}{3}$, and $\alpha_{12} = 109\,°\,28\,'$ while $-\frac{1}{\mu^2} = -\frac{1}{3}$, and $\mu = \sqrt{3}$. From the equation $\Psi_1 = \alpha_1 (S + \mu_1\, \zeta_1)$, it is possible to write that $\Psi_1 = \alpha_1 (S + \sqrt{3}\, \zeta_1)$. The normalization condition for Ψ_1 gives a value for α_1: $< \Psi_1 \mid \Psi_1 > = 1 = \alpha_1^2\, [1 + (\sqrt{3})^2] = 4\, \alpha_1^2$, so that $\alpha_1 = \frac{1}{2}$. We find that: $\Psi_1 = \frac{1}{2}\, (S + \sqrt{3}\, \zeta_1) = \frac{1}{2}\, [S + (a'_1\, X + b'_1\, Y + c'_1\, Z)\sqrt{3}\,]$, where (a'_1, b'_1, c'_1) are the cosine directors for the direction of the first orbital that has been chosen such that $a'_1 = b'_1 = c'_1 = \frac{1}{\sqrt{3}}$. This gives:

$$\Psi_1 = \frac{1}{2}\left(\mathrm{S} + \sqrt{3}\left(\frac{1}{\sqrt{3}}\varphi_{2p_x} + \frac{1}{\sqrt{3}}\varphi_{2p_y} + \frac{1}{\sqrt{3}}\varphi_{2p_z}\right)\right).$$

At the end of all this, we finally obtain:

$$\begin{cases} \Psi_1 = \dfrac{1}{2}\,(S + \varphi_{2p_x} + \varphi_{2p_y} + \varphi_{2p_z}) = \dfrac{1}{2}(S + X + Y + Z) \\ \Psi_2 = \dfrac{1}{2}\,(S - \varphi_{2p_x} - \varphi_{2p_y} + \varphi_{2p_z}) = \dfrac{1}{2}\,(S - X - Y + Z) \\ \Psi_3 = \dfrac{1}{2}\,(S + \varphi_{2p_x} - \varphi_{2p_y} - \varphi_{2p_z}) = \dfrac{1}{2}\,(S + X - Y - Z) \\ \Psi_4 = \dfrac{1}{2}\,(S - \varphi_{2p_x} + \varphi_{2p_y} - \varphi_{2p_z}) = \dfrac{1}{2}\,(S - X + Y - Z) \end{cases}$$



As observed at the end of section 8.2.3.1, the Ψ_i functions revolve around $\vec{t}_i^{\,'}$. It must be noted that these functions are directed towards the summit of the tetrahedron as shown in a covalent bond.

8.3. Molecular model of a 3D covalent crystal (atoms in sp^3-hybridization states at lattice nodes)

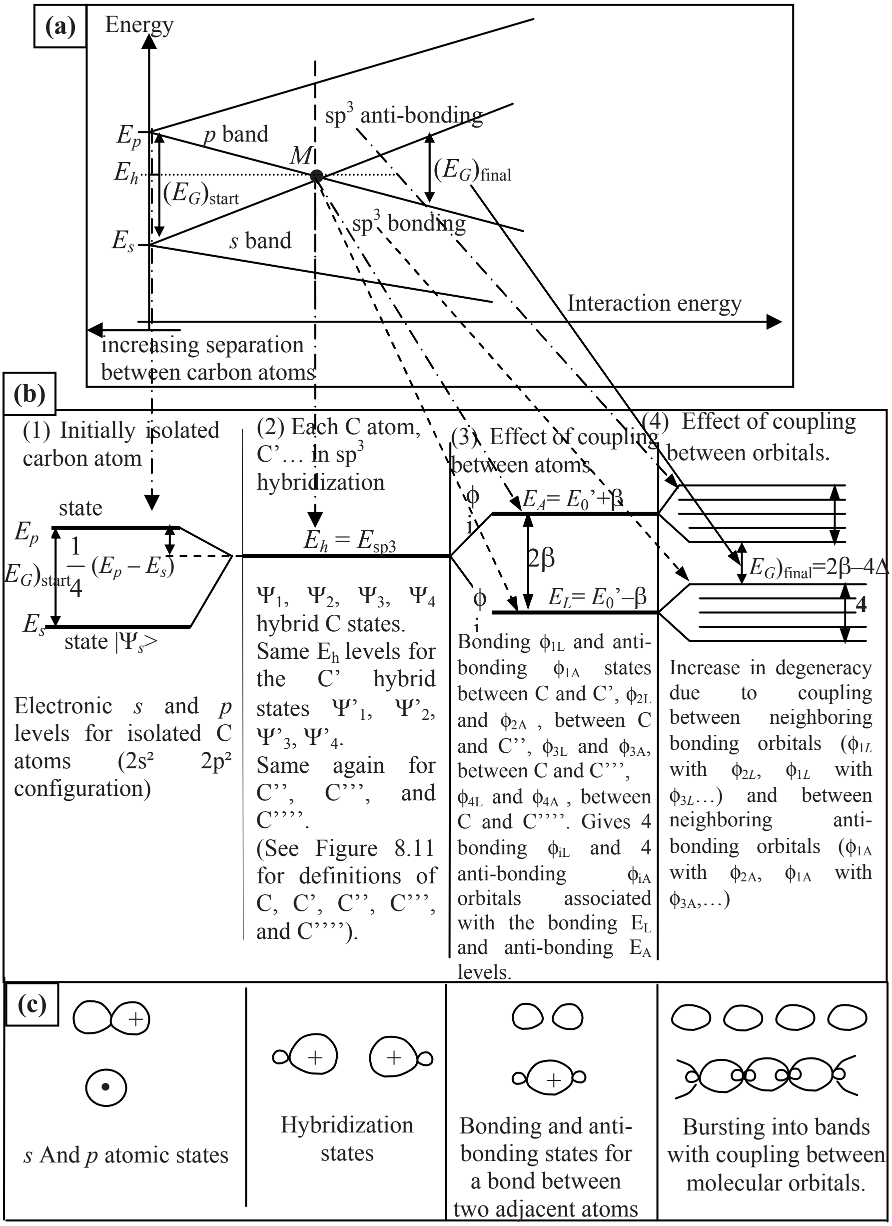
8.3.1. *Conditions*

The carbon atoms in diamond exhibit a *sp*^{3}-hybridization. The various forms that carbon orbitals take in 3D space are detailed in the preceding section. If the atomic orbitals are effectively equally spread in different directions, then we have a *sp*^{3}-hybridization. And in this case, using the carbon atom denoted C as reference, then the orbitals are expressed using the four functions that have previously been calculated, namely $|\Psi_1>$, $|\Psi_2>$, $|\Psi_3>$, $|\Psi_4>$.

Figure 8.10 details the various stages during the formation of the electronic states and energy levels in diamond.

8.3.1.1. *It all starts with isolated carbon atoms*

Initially, the isolated carbon atoms are characterized by two energy levels, E_s and E_p, for the electronic configuration $1s^2\ 2s^2\ 2p^2$. This is shown in Figure 8.10a and in zone (1) of Figure 8.10b. The atoms *C*, *C*', *C*''… shown in Figure 8.11 are at this stage presumed to be separated by a great distance. As the atoms *C*', *C*'', *C*''' and *C*'''' move nearer to the reference atom *C*, there are *s* and *p* bands that form, as shown in Figure 8.10a. This is concurrent with the overlapping of the orbitals of each atom, for example the *s*-orbitals give bonding combinations towards the bottom and anti-bonding combinations towards the top (see Chapter 7).



Electronic s and p levels for isolated C atoms (2s² 2p² configuration)

Ψ_1, Ψ_2, Ψ_3, Ψ_4 hybrid C states. Same E_h levels for the C' hybrid states Ψ'_1, Ψ'_2, Ψ'_3, Ψ'_4. Same again for C'', C''', and C''''. (See Figure 8.11 for definitions of C, C', C'', C''', and C'''').

Bonding ϕ_{1L} and anti-bonding ϕ_{1A} states between C and C', ϕ_{2L} and ϕ_{2A}, between C and C'', ϕ_{3L} and ϕ_{3A}, between C and C''', ϕ_{4L} and ϕ_{4A}, between C and C''''. Gives 4 bonding ϕ_{iL} and 4 anti-bonding ϕ_{iA} orbitals associated with the bonding E_L and anti-bonding E_A levels.

Increase in degeneracy due to coupling between neighboring bonding orbitals (ϕ_{1L} with ϕ_{2L}, ϕ_{1L} with ϕ_{3L}…) and between neighboring anti-bonding orbitals (ϕ_{1A} with ϕ_{2A}, ϕ_{1A} with ϕ_{3A},…)

Figure 8.10. *(a) Formation of bands with coalescing carbon atoms; (b) evolution of the electronic energy levels with successive couplings; and (c) schematic illustration of orbital evolution for the associated changes in energy levels*

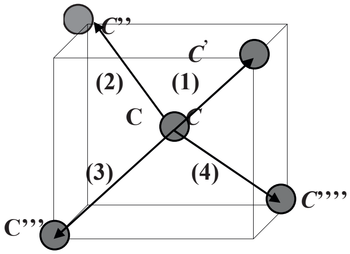


Figure 8.11. *Layout for the different atoms along with the possible couplings using sp^3-hybridization*

8.3.1.2. *At the critical point M*

The two bands, *s* and *p*, meet at the critical point *M* and the *s* and *p* states mix (the so-called process of hybridization). With this hybridization being energetically favorable, we obtain *sp*^{3}-hybrid states denoted by the Ψ_1, Ψ_2, Ψ_3, Ψ_4 functions for the atom *C* (see section 8.2). Similarly, the hybrid states for the atom *C*' are represented by the functions Ψ'_1, Ψ'_2, Ψ'_3, Ψ'_4, and so on in similar terms of notation for the *C*'', *C*''' and *C*'''' atoms (for example the sp^3-states of the last atom are described by the functions Ψ_1'''', Ψ_2'''', Ψ_3'''', Ψ_4''''). Each group of states, shown in zone (2) of Figure 8.10b, are represented by the functions $\Psi_{i)}{}_{i\,=\,1,2,3,4}$, $\Psi'_{i)}{}_{i\,=\,1,2,3,4}$ and so on (so that there are 4*N* states if the system contains *N* atoms) with the same energy levels, as in $E_{\mathrm{sp3}} = E_h$.

E_h can be calculated simply as, for example:

$$E_h = <\Psi_1|\, H \,|\, \Psi_1> = <\frac{1}{2}\ (S + \varphi_{2p_x} + \varphi_{2p_y} + \varphi_{2p_z}\ )|H|\frac{1}{2}$$

$$\times\ (S + \varphi_{2p_x} + \varphi_{2p_y} + \varphi_{2p_z}\ )$$

$$= \frac{1}{4}\ \{<S\,|\mathrm{H}|\ S> + <\varphi_x\ |H|\ \varphi_x > + <\varphi_y\ |H|\ \varphi_y> + <\varphi_z\ |H|\ \varphi_z>$$

$$= \frac{1}{4}\ \{E_s + 3E_p\},$$

where also:

$$E_h = <\Psi_2|\, H \,|\, \Psi_2> = <\Psi_3\ H \,|\, \Psi_3> = <\Psi_4|\, H \,|\, \Psi_4> = <\Psi'_1|\, H \,|\, \Psi'_1> = E'_0.$$

(E_p and E_s represent, respectively, the energy levels of the 2*p* and 2*s* states; see zone (1) of Figure 8.10b).

We thus find that :

$$E_p - E_h = E_p - \frac{1}{4}\{E_s + 3\,E_p\} = \frac{1}{4}(E_p - E_s). \qquad [8.12]$$

8.3.1.3. *Initial follow-up study of the effect of type A coupling between nearest neighbor carbon atoms*

This section is limited to dealing with coupling between the bonds denoted (1), (2), (3), and (4), as identified in Figure 8.12. Put in other terms, it is limited to looking at coupling effects between the bonds denoted: C–C’, C–C’’, C–C’’’, and C–C’’’’, without considering the effects of other bonds between those atoms and their other nearest neighbors. There is a formation of bonding and anti-bonding states, qualitatively described in zone (3) of Figure 8.10b, and more quantitatively detailed in section 8.3.2.

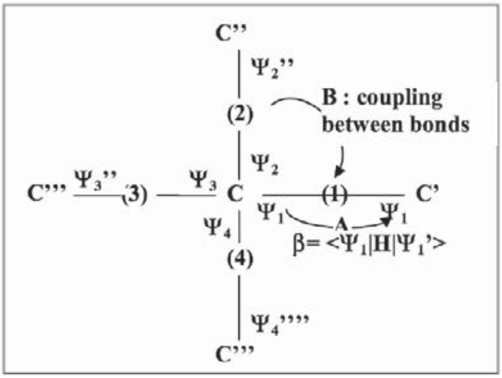


Figure 8.12. *Representation of A and B couplings considered successively (obtained as a projection from Figure 8.11)*

8.3.1.4. *Initial study of the supplementary effect of B-type coupling (see Figure 8.12) between molecular orbitals*

The coupling between (1) and (2), and between (2) and (3), and so on, that generates energy bands is shown in zone (4) in Figure 8.10b. Once again, a more quantitative approach will be developed in section 8.3.3.

8.3.2. *Independent bonds: effect of single coupling between neighboring atoms and formation of molecular orbitals*

Zone (3) corresponds to the states appearing following paired coupling between two atoms using the sp^3 hybridization (for example between atoms C and C’). This atomic orbital coupling result in a formation of molecular orbitals for which there are Φ solutions that can be given in the form of a linear combination of atomic orbitals (with the $|\Psi>$ orbital for the C atom, and the $|\Psi'>$ orbital for the C’ atom), i.e.:

$$\Phi = c\,|\Psi> + c'|\Psi'>.$$

If the direction denoted 1 represents the bond between atoms C and C’, the resulting molecular orbital (Φ_1) can be bonding or anti-bonding (see sections 7.1.2 and 7.1.3), as in:

$$\Phi_{1L} = \frac{1}{\sqrt{2}}(|\Psi_1> + |\Psi_1'>) \qquad [8.13]$$

$$\Phi_{1A} = \frac{1}{\sqrt{2}}(|\Psi_1> - |\Psi_1'>)$$

With the pairs being tied to one single carbon atom but nevertheless independent, that is to say without affecting one another, there appear (in addition to Φ_{1L} and Φ_{1A}) the following molecular orbitals around the atom C:

$$\Phi_{2L} = \frac{1}{\sqrt{2}}(|\Psi_2> + |\Psi_2''>) \quad \text{and} \quad \Phi_{2A} = \frac{1}{\sqrt{2}}(|\Psi_2> - |\Psi_2''>)$$

$$\Phi_{3L} = \frac{1}{\sqrt{2}}(|\Psi_3> + |\Psi_3'''>) \quad \text{and} \quad \Phi_{3A} = \frac{1}{\sqrt{2}}(|\Psi_2> - |\Psi_3'''>)$$

$$\Phi_{4L} = \frac{1}{\sqrt{2}}(|\Psi_2> + |\Psi_4''''>) \quad \text{and} \quad \Phi_{4A} = \frac{1}{\sqrt{2}}(|\Psi_2> - |\Psi_4''''>)$$

The energies associated with these bonding and anti-bonding states are the respective energies E_L and E_A. They have the same form as those determined in section 7.1.2, as in:

$$E_L = E'_0 - \beta \quad \text{and} \quad E_A = E'_0 + \beta \qquad [8.14]$$

where:

$$E'_0 = H_{ii} = <\Psi_i\,|H|\Psi_i> = E_h = E_{sp3}$$

$$-\beta = H_{ii'} = <\Psi_i\,|H|\Psi_i'> \quad \text{(coupling parameter between paired near-neighbors)}$$

NOTE.– With N being the number of atoms in a crystal, the number of Φ_L-type bonds is equal to $2N$ as each carbon atom gives rise to four bonding bonds (each containing two electrons) with each being shared between two atoms (therefore on average each carbon atom has two bonding bonds).

With the number of valence electrons per atom being four ($2s^2\ 2p^2 \rightarrow 2t^4$, with t representing the hybrid states), the fundamental state corresponds to the placement of $4\times N$ electrons, and this can also be written as ($2\times 2N$) as it represents the number of electrons per bonding orbital multiplied by the number of bonding bonds.

The consequence of this is that all bonding bonds are full and the anti-bonding bonds (Φ_A, that are present in an equal number) are all empty.

8.3.3. *Coupling of molecular orbitals: band formation*

In effect, the reasoning in section 8.3.2 highlights that for a crystal of N atoms the level denoted E_L (just like E_A) is degenerated $2N$ times (as it is for $2N$ bonding orbitals). The following discussion centers on the effect of coupling between orbitals on their degeneracy.

8.3.3.1. *Effect of the bonding energy between hybrid orbitals around the same carbon atom*

This energy can be stated as $<\Psi_1\,|\,H\,|\,\Psi_2> = -\Delta$, so that with the forms given for Ψ_1 and Ψ_2 in section 8.2,

$$-\Delta = <\frac{1}{2}(S+X+Y+Z\,|\,H\,|\,\frac{1}{2}(S-X-Y+Z)>$$

$$= \frac{1}{4}(E_s - E_p - E_p + E_p) = \frac{1}{4}(E_s - E_p)$$

The effect is apparently non-zero, meaning that in a 3D covalent crystal we should expect a non-zero coupling between the molecular orbitals bonding two adjacent atoms.

8.3.3.2. *Effect of coupling between neighboring bonding orbitals in the heart of the crystal*

This coupling, denoted by the letter B in Figure 8.12, has matrix elements in the form:

$$<\Phi_{1L} \mid H \mid \Phi_{2L}> = <\frac{1}{\sqrt{2}}(|\Psi_1> + |\Psi_1'>) \mid H \mid \frac{1}{\sqrt{2}}(|\Psi_2> + |\Psi_2''>)>$$

Neglecting the coupling integrals between non-adjacent neighbors (so that for example, $<\Psi_1'| H |\Psi_2> \approx 0$), we obtain:

$$<\Phi_{1L} \mid H \mid \Phi_{2L}> = \frac{1}{2}<\Psi_1 \mid H \mid \Psi_2> = \frac{1}{2} \times \frac{1}{4}(E_s - E_p) = -\frac{\Delta}{2} \qquad [8.15]$$

By modifying the energy levels (type E_L), the coupling of molecular bonds increases the degeneration by $E_L = E'_0 - \beta$.

By analogy to the 1D model, the wave functions of the crystal need to be written, at this level, as a linear combination of bonding (Φ_L) or anti-bonding orbitals (Φ_A). These functions, characteristic of a periodic lattice, should always satisfy the Bloch and Floquet theorems, and therefore should take on the form:

$$\left|\Phi_k^L(\vec{r})\right\rangle = c_0 \sum_s e^{i\vec{k}.\vec{r}_s} \left|\Phi_L(\vec{r} - \vec{r}_s)\right\rangle$$

$$\left|\Phi_k^A(\vec{r})\right\rangle = c'_0 \sum_s e^{i\vec{k}.\vec{r}_s} \left|\Phi_A(\vec{r} - \vec{r}_s)\right\rangle$$

These are the additive Bloch functions for bonding and anti-bonding functions. They delocalize the electrons throughout the 3D lattice, just as the equations verifying Floquet's theorem do in a 1D system (see equation [7.16]).

Just as in the 1D systems detailed in Chapter 7 and the 3D system based on *s*-orbitals described in this Chapter 8 (section 8.1), these functions result in the approximate levels E_L and E_A (zone 3 of Figure 8.10) breaking down into bands (zone 4 of Figure 8.10).

This gives rise to $2N$ functions of the type $\left|\Phi_k^L(\vec{r})\right\rangle$ (in fact $4N$ if you take spin into account). The bond of bonding states is therefore full, justifying the terminology "highest occupied molecular orbital" or HOMO for short, otherwise known to physicists as the valence band (VB). For its part, the band of anti-bonding states, known as the lowest unoccupied molecular orbital or LUMO, and conduction band (CB) is empty. These two bands are separated by the so-called "gap" that has a height denoted E_G.

In quantitative terms, we have seen in section 8.1 that for *s*-orbitals characterized by the coupling parameter $-\beta = <\psi_s|H|\psi_{s\pm1}>$), the width of the thus formed bands I equal to $2Z\beta$ in the alternating systems (see equations [8.10]). For the example treated in Figure 8.11, the coordination number of Z is 4 and the coupling parameter is given by $<\Phi_{1L} \mid H \mid \Phi_{2L}> = -\frac{\Delta}{2}$. An estimation of the width of the HOMO and LUMO bands can thus be given by assuming that equation [8.10] is acceptable for the orbitals is acceptable, so that $B \simeq 2.4.(|\Delta|/2) = 4\,|\Delta|$ (zone 4 in Figure 8.10b).

In addition, the height of the gap can be directly evaluated from Figure 8.10b where $E_G \simeq 2\beta - 4|\Delta|$. The values of β and Δ depend on the cell parameters and the sizes of the atoms. For example, with the gap E_G = 5.4 eV, diamond is more of an insulator than a semiconductor.

On going to the bottom of column IV in the periodic table, that is from carbon through silicon towards germanium, i.e. with increasing atom size, there is an increase (≈ 4Δ) in the permitted bands, so that the gap decreases (5.4 eV for C, 1.1 eV for Si, 0.7 eV for Ge).

8.4. Complementary in-depth study: determination of the silicon band structure using the strong bond method

This study comes from the work of G. Leman in "*Annales de physique*", vol. 7, pages 505–533 (1962). Given the complexity of the calculations, this work is considered as a complementary study to the simplified version shown in section 8.3 that deals with diamond, that has the same structure as silicon.

8.4.1. *Atomic wave functions and structures*

The structure of silicon is similar to that of the carbon diamond; it is based on an inter-penetration of two fcc lattices, one being displaced with respect to the other by a quarter diagonal (see Figures 6.6b, and 8.4). The atoms represented by an empty circle, those of the second fcc, in Figures 8.4 and 8.13 are positioned by the vector $\vec{r}_j^{\,'}$ shifted by a vector $\vec{b}$ with components $\left(\frac{d}{4},\frac{d}{4},\frac{d}{4}\right)$ with respect to the $\vec{r}_j$ vector, which locates the full circled atoms of the first fcc. In effect, $\vec{r}_j^{\,'} = \vec{r}_j + \vec{b}$. The vectors are defined as follows so that we have a systematic and coherent set of notation:

$$\vec{a}_0\begin{cases}0\\0,\\0\end{cases}\quad \vec{a}_1\begin{cases}d/2\\d/2,\\0\end{cases}\quad \vec{a}_2\begin{cases}d/2\\0\quad ,\\d/2\end{cases}\quad \vec{a}_3\begin{cases}0\\d/2\\d/2\end{cases}$$

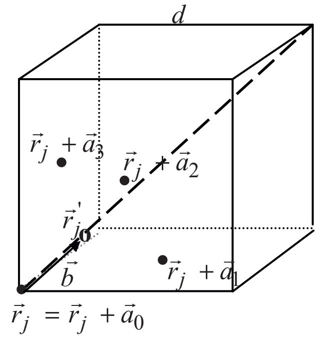


Figure 8.13. *Location of atoms in the silicon based structure*

The first four atoms neighboring the atom located by $\vec{r}_j^{\,'}$ (belonging to the second lattice) are located by the vectors $\vec{r}_j + \vec{a}_n$. Similarly, the first four atoms neighboring the atom located by $\vec{r}_j$ (belonging to the first lattice) are located by the vectors $\vec{r}_j^{\,'} - \vec{a}_n$. Each atom located by $\vec{r}_j$ (or by $\vec{r}_j^{\,'}$) is at the center of a regular tetrahedron, while the atomic orbitals that bond them to their closest neighbors are hybrid orbitals with a form determined in section 8.2.3.3 and denoted Ψ_1, Ψ_2, Ψ_3, Ψ_4 (the electronic configuration of the external layer of silicon is given by $3s^2 3p^2$, so the Ψ functions are obtained from the φ_{3p} orbitals instead of the φ_{2p} orbitals used in the case of carbon). Here, the hybrid functions for the atoms located by $\vec{r}_j$ or $\vec{r}_j^{\,'}$ are denoted using φ_{ji} and $\varphi_{ji}^{'}$, with $i = 0, 1, 2, 3$, such that:

$$\left\{\begin{array}{ll}
\varphi_{j0} = \frac{1}{2}\,(S + \varphi_{3p_x} + \varphi_{3p_y} + \varphi_{3p_z}) & \text{goes in the direction [111]} \\
\varphi_{j1} = \frac{1}{2}\,(S - \varphi_{3p_x} - \varphi_{3p_y} + \varphi_{3p_z}) & \text{goes in the direction [-1-11]} \\
\varphi_{j2} = \frac{1}{2}\,(S + \varphi_{3p_x} - \varphi_{3p_y} - \varphi_{3p_z}) & \text{goes in the direction [1-1-1]}
\end{array}\right.$$

$$\left\{\begin{array}{ll}
\varphi_{j3} = \frac{1}{2}\,(S - \varphi_{3p_x} + \varphi_{3p_y} - \varphi_{3p_z}) & \text{goes in the direction [-11-1]} \\
\varphi'_{j0} = \frac{1}{2}\,(S' - \varphi'_{3p_x} - \varphi'_{3p_y} - \varphi'_{3p_z}) & \text{goes in the direction [-1-1-1]} \\
\varphi'_{j1} = \frac{1}{2}\,(S' + \varphi'_{3p_x} + \varphi'_{3p_y} - \varphi'_{3p_z}) & \text{goes in the direction [11-1]} \\
\varphi'_{j2} = \frac{1}{2}\,(S' - \varphi'_{3p_x} + \varphi'_{3p_y} + \varphi'_{3p_z}) & \text{goes in the direction [-111]} \\
\varphi'_{j3} = \frac{1}{2}\,(S' + \varphi'_{3p_x} - \varphi'_{3p_y} + \varphi'_{3p_z}) & \text{goes in the direction [1-11]}
\end{array}\right.$$

These functions (atomic orbitals) make up a useful base for describing tetrahedral bonds, but they are not the proper functions of a crystalline system (orbitals described by following the LCAO method). So, each function points towards a tetrahedral summit, the identity of which depends on the identity of the covalent bond being studied. Schematically, for a bond between an atom of the first lattice located by $\vec{r}_j$ (= $\vec{r}_j + \vec{a}_0$) and an atom of the second lattice located by $\vec{r}_j^{\,'}$ ($= \vec{r}_j^{\,'} - \vec{a}_0$), Figure 8.14 shows the line along which runs the direction of the bond.

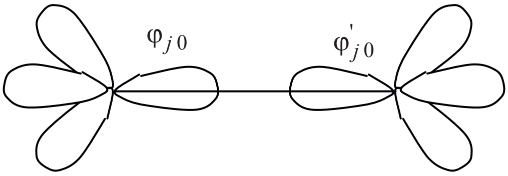


Figure 8.14. *Bond between atoms located by* $\vec{r}_j$ *and* $\vec{r}_j^{\,'}$

If H represents the Hamiltonian for an atom, of form $H = -\frac{\hbar^2}{2m}\Delta + V(\vec{r} - \vec{r}_j)$ for the atom centered on $\vec{r}_j$, we can use the calculations in sections 8.3.1.2 and 8.3.3.1, respectively:

– a matrix cell with a form given by:

$$\begin{aligned}\left\langle \varphi_{j0} \mid H \mid \varphi_{j0} \right\rangle &= \frac{1}{4}\left\{E_s + 3E_p\right\}\\ &= \left\langle \varphi_{ji} \mid H \mid \varphi_{ji} \right\rangle\end{aligned}$$

– a matrix cell with a form given by:

$$\begin{aligned}\left\langle \varphi_{j0} \mid H \mid \varphi_{j1} \right\rangle = -\Delta &= \frac{1}{4}\left(E_s - E_p\right)\\ &= \left\langle \varphi_{ji} \mid H \mid \varphi_{j\ell} \right\rangle, \text{ with } i \neq \ell\end{aligned}$$

Evidently this gives the same results as those with the φ'_{ji} functions.

8.4.2. *Wave functions in crystals and equations with proper values for a strong bond approximation*

8.4.2.1. *Wave functions and their properties*

For a crystal and point located by $\vec{r}$, the Hamiltonian is in the form:

$$H_{\text{cryst}} = -\frac{\hbar^2}{2m}\Delta + \sum_j \left[V(\vec{r} - \vec{r}_j) + V(\vec{r} - \vec{r}'_j)\right],$$

where $V(\vec{r} - \vec{r}_j)$ and $V(\vec{r} - \vec{r}'_j)$ are the Coulombic potentials centered at atoms $\vec{r}_j$ and r'_j, respectively.

The wave function at a point $\vec{r}$ in the crystal, within a strong bond approximation, must appear as a linear combination of atomic orbitals for which there is a limit in the number of cells directed towards the first four neighbors.

Similarly, this function should have in front of it a Bloch function and the final solution should in the (Floquet) form

$$\Psi_k\left(\vec{r}\right) = C\sum_j e^{i\vec{k}.\vec{r}_j} \sum_{i=0,1,2,3}\left[A_i\,\varphi_{ji}\,(\vec{r}-\vec{r}_j) + A_i^{'}\,\varphi_{ji}\,(\vec{r}-\vec{r}_j^{\,'})\right]$$

where there is a sum over all sites (*j*) in the lattice, the index *i* represents one of four orbitals at a *j* site and *C* is the normalization factor. In the simplified model developed here, all φ_{ji} or $\varphi_{ji}^{'}$ *pseudo*-orthogonal functions are considered, while neglecting the overlap integrals (before the unit) of the type $\left\langle \varphi_{ji} \mid \varphi_{\ell m} \right\rangle_{j\neq\ell} = 0$ which is the overlap integral for orbitals in the same fcc lattice centered on different atoms ($j \neq \ell$); and $\left\langle \varphi_{ji} \mid \varphi_{\ell m}^{'} \right\rangle = 0$ which is the overlap integral between orbitals from different lattices (for whatever ℓ, even $\ell = j$).

The only overlap integrals between atomic centers taken into account are those for immediate neighbors (that also therefore belong to two different fcc lattices) and only their atomic orbitals that point directly towards one another (Hückel type approximation). The only integrals that are studied are therefore given by (wherein j and ℓ denote the first neighbors):

$$\left\langle \varphi_{ji} \;|V\,(\vec{r}-\vec{r}_j)\,|\, \varphi_{li}^{'} \right\rangle = \left\langle \varphi_{ji} \;|V\,(\vec{r}-\vec{r}_\ell^{\,'})\,|\, \varphi_{li}^{'} \right\rangle = -\lambda$$

Thus $\vec{r}_\ell^{\,'} = \vec{r}_j - \vec{a}_n$ locates the first neighbor to that at $\vec{r}_j$ towards which points the φ_{ji} function. The parameter λ is positive because the potentials (V) are negative (attracting potentials). These hypotheses of course neglect completely any influence of second neighbors. So to resume, we can use $\varphi_{\rightarrow}$ and $\varphi_{\leftarrow}$ to denote the functions on neighboring atoms that point towards one another so that the strong bond approximation and the overlap integrals between immediate neighbors (Hückel) can be written, respectively as

$$\left\langle \varphi_{\rightarrow} \mid \varphi_{\leftarrow}^{'} \right\rangle = 0 \text{ and } \left\langle \varphi_{\rightarrow} \mid V \;\mid \varphi_{\leftarrow}^{'} \right\rangle = \left\langle \varphi_{\rightarrow} \mid V\,' \mid \varphi_{\leftarrow}^{'} \right\rangle = -\lambda.$$

8.4.2.2. *Equation using proper values*

To simplify the problem, we can consider just the atom at the origin. If H is the Hamiltonian of the electron placed at the central atom, the Hamiltonian for the crystal can be:

$$H_{\text{cryst}} = H + \sum_{j \neq 0} V(\vec{r} - \vec{r}_j) + \sum_{j} V(\vec{r} - \vec{r}_j^{\,'})$$

Using φ_{0i} and $\varphi_{0i}^{'}$ as the hybrid functions for, respectively, the atom at the origin and the atom situated with respect to $\vec{b}$ by components $1/4, 1/4, 1/4$, we can perform a scalar multiplication on the left-hand side successively with $\langle \varphi_{0i} |$ and $\langle \varphi_{0i}^{'} |$ to give the equation with proper values:

$$H_{\text{cryst}} \left| \psi_k(\vec{r}) \right\rangle = E_k \left| \psi_k(\vec{r}) \right\rangle$$

and obtain the two equations (with $i = 0, 1, 2, 3$):

$$\left\langle \varphi_{0i} \right| H_{\text{cryst}} \left| \psi_k(\vec{r}) \right\rangle = E_k \left\langle \varphi_{0i} \mid \psi_k(\vec{r}) \right\rangle$$

$$\left\langle \varphi_{0i}\,' \right| H_{\text{cryst}} \left| \psi_k(\vec{r}) \right\rangle = E_k \left\langle \varphi_{0i}^{'} \mid \psi_k(\vec{r}) \right\rangle.$$

At the origin, the atomic orbital (φ_{00}) points towards a site placed along $\vec{b}$ found (according to our simplified hypothesis) uniquely coupled with the orbital $\varphi_{00}^{'}$, and with a coupling coefficient given by $e^{i\vec{k}.\vec{a}_0} \overset{a_0=0}{=} e^{0} = 1$. Similarly, at the origin, the φ_{01} orbital directed along [-1-11] is only coupled with the orbital centered on the site with coordinates ($-a/4$, $-a/4$, $a/4$), vector components $\vec{r}_j^{\,'} = \vec{b} - \vec{a}$, and points towards [11-1], so that we have the orbital given by $\varphi_{-1,1}^{'}(\vec{r} - [-\vec{a}_1]) = \varphi_{-1,1}^{'}(\vec{r} - \vec{a}_1)$. As $\vec{r}_j^{\,'} = \vec{r}_j + \vec{b}$, we here have $\vec{r}_j = -\vec{a}_1$, and the corresponding coupling coefficient (in the form $e^{i\vec{k}.\vec{r}_j}$) is $e^{-i\vec{k}.\vec{a}_1}$. By extending these results over all φ_{0i} orbitals, the two preceding equations arising from the proper values equation then give:

$$\left(\frac{E_s+3E_p}{4}-E_k\right)A_m+\left(\frac{E_s-E_p}{4}\right)\sum_{i\neq m}A_i-\lambda e^{-i\vec{k}.\vec{a}_m}A'_m=0$$

$$-\lambda e^{i\vec{k}.\vec{a}_m}A_m+\left(\frac{E_s+3E_p}{4}-E_k\right)A'_m+\left(\frac{E_s-E_p}{4}\right)\sum_{i\neq m}A'_i=0$$

By making $x=(E_k-E_p)$, $\alpha_n=e^{-i\vec{k}\cdot\vec{a}_n}$, and as stated above $\Delta=\frac{E_p-E_s}{4}$, the preceding system can be written as:

$$\left(x+\frac{\Delta}{2}\right)A_m+\frac{\Delta}{2}\sum_{i\neq m}A_i-\lambda\alpha_m A'_m=0$$

$$-\lambda\alpha_m^* A_m+\left(x+\frac{\Delta}{2}\right)A'_m+\frac{\Delta}{2}\sum_{i\neq m}A'_i=0.$$

This system based on $2\times 4=8$ linear equations (as i takes on the four values 0, 1, 2, 3) gives eight coefficients (A_i and A'_i) from which the wave function $\Psi_k(r)$ can be deduced. In order to do this, we shall write that the compatibility of the eight equations necessitates that the determinant of the system is equal to zero, hence we have the following secular equation:

$$\begin{vmatrix}
x+\Delta & \Delta & \Delta & \Delta & \lambda\alpha_0 & 0 & 0 & 0\\
\Delta & x+\Delta & \Delta & \Delta & 0 & \lambda\alpha_1 & 0 & 0\\
\Delta & \Delta & x+\Delta & \Delta & 0 & 0 & \lambda\alpha_2 & 0\\
\Delta & \Delta & \Delta & x+\Delta & 0 & 0 & 0 & \lambda\alpha_3\\
\lambda\alpha_0^* & 0 & 0 & 0 & x+\Delta & \Delta & \Delta & \Delta\\
0 & \lambda\alpha_1^* & 0 & 0 & \Delta & x+\Delta & \Delta & \Delta\\
0 & 0 & \lambda\alpha_2^* & 0 & \Delta & \Delta & x+\Delta & \Delta\\
0 & 0 & 0 & \lambda\alpha_3^* & \Delta & \Delta & \Delta & x+\Delta
\end{vmatrix}=0.$$

By making $\phi=\frac{1}{4}\sum_n\alpha_n$, a long calculation shows that the secular equation can be written as:

$$\left(x^2-\lambda^2\right)^2\left(x^2+4x\Delta-\lambda^2+4\lambda\phi\Delta\right)\left(x^2+4x\Delta-\lambda^2-4\lambda\phi\Delta\right)=0$$

where ϕ is a Bloch function that depends on the $\vec{k}$ vector and is given by

$$\phi(\vec{k}) = \frac{1}{2}\left[1 + \cos\frac{k_x a}{2}\cos\frac{k_y a}{2} + \cos\frac{k_x a}{2}\cos\frac{k_z a}{2} + \cos\frac{k_y a}{2}\cos\frac{k_z a}{2}\right]^{1/2} = 0$$

When $\vec{k}$ is directed along [100], such that $k_x = k$ and $\mathrm{k_y} = k_z = 0$, we thus have $\phi = \cos\dfrac{ka}{4}$.

8.4.3. *Band structure*

$x\left(\phi\right)$ can be determined using the secular equation, and then we can find the dispersion relation $E(\vec{k})$. As ϕ depends on the direction of $\vec{k}$, the energy $E_k = E_p + x\left(\phi\right)$ also depends on the direction of $\vec{k}$. The equi-energy surfaces in the reciprocal space are undulated spheres, undergoing so-called "warping". The energy E ($\equiv E_k$) is in fact a function of two parameters, on the one hand $\Delta = \frac{1}{4}(E_p - E_s)$ which depends on the nature of the free atom, but also varies little with each type of element (between C, Si, and Ge), and on the other hand, the overlap integral (λ) for two neighbors that depends on the parameter of the crystalline lattice (*d*), which is in fact a function that decreases with the distance (δ) between atoms. The energy is generally studied as a function of $2\Delta/\lambda$ and we can show that the electronic structure changes considerably depending on whether or not λ is higher or lower than 2Δ.

8.4.3.1. *Flat bands*

These relate to the solution found for the secular equation when:

$$\left(x^2 - \lambda^2\right)^2 = 0$$

in that:

$$x_1 = E_1 - E_p = +\lambda \text{ (twice degenerate solution)}$$

$$x_2 = E_2 - E_p = -\lambda \text{ (twice degenerate solution).}$$

These bands are flat in the sense that the values of E_1 and E_2 of the energy does not depend on *k*.

8.4.3.2. *Wide bands*

These are associated with solutions other than those of the secular equation, and correspond:

– on the one hand to $\left(x^2 + 4x\,\Delta - \lambda^2 - 4\lambda\phi\Delta\right) = 0$, with solutions:

$$x_3 = -2\Delta + \sqrt{4\Delta^2 + \lambda^2 + 4\lambda\phi\Delta}$$

$$x_4 = -2\Delta - \sqrt{4\Delta^2 + \lambda^2 + 4\lambda\phi\Delta}$$

– on the other hand to $\left(x^2 + 4x\,\Delta - \lambda^2 + 4\lambda\phi\Delta\right) = 0$, with solutions:

$$x_5 = -2\Delta + \sqrt{4\Delta^2 + \lambda^2 - 4\lambda\phi\Delta}$$

$$x_6 = -2\Delta - \sqrt{4\Delta^2 + \lambda^2 - 4\lambda\phi\Delta}$$

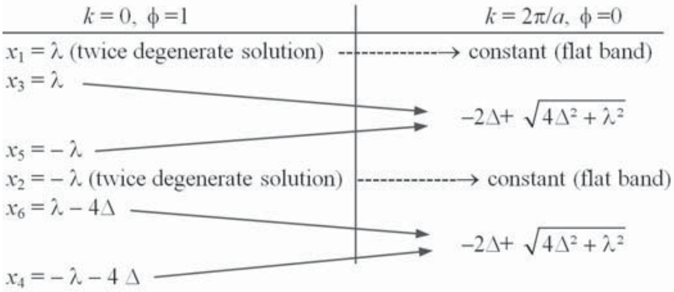
In the direction [100] where $\phi = \cos\frac{ka}{4}$, when k varies from 0 to $\frac{2\pi}{a}$, ϕ changes from 1 to 0 and the energies are given as solutions x_i (i = 3,4,5,6) that vary continuously from $x_i\,(0)$ to $x_i\left(\frac{2\pi}{a}\right)$, leading to wide bands.

With $k = 0$, we have $x_5 = -\Delta + \left|2\Delta - \lambda\right|$ and $x_6 = -\Delta - \left|2\Delta - \lambda\right|$, which shows that when $\lambda = 2\Delta$ we have a change in behavior.

8.4.3.3. *Practical study of E = f(k), or more exactly x = E – E$_p$ as a function of k when k is parallel to the direction [100] and for which* $\phi = \cos(ka / 4)$

Over the width of the zone, k varies from 0 to $2\pi / a$ and ϕ respectively takes on 1 and 0.

When $\lambda < 2\Delta$, the solutions for x, with respect to k values, are the following (see also Figure 8.15a):



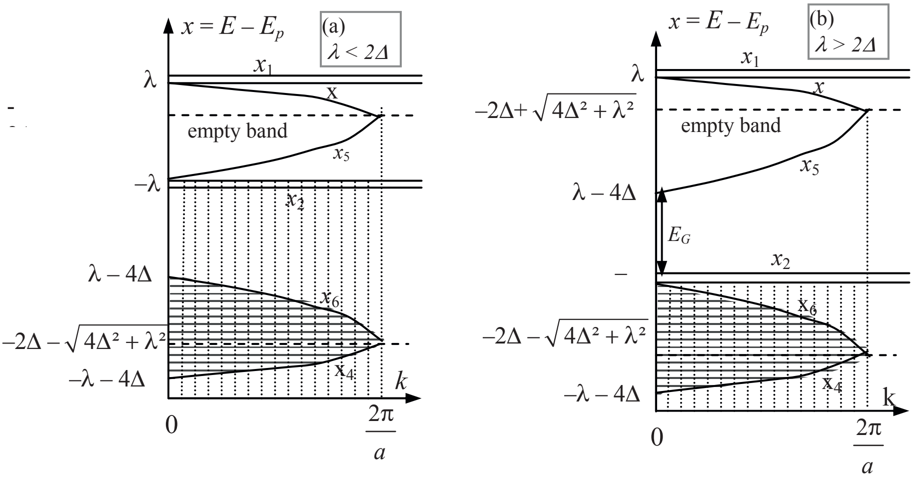
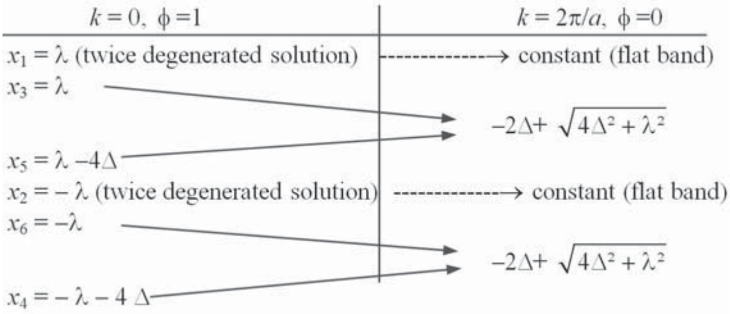


Figure 8.15. *Curves for $x = f(k)$ when: (a) $\lambda < 2\Delta$ which corresponds to having a large δ; and (b) $\lambda > 2\Delta$ as when δ is small*

When $\lambda > 2\Delta$, the solutions for x, depending on k values, are (also see Figure 8.15b):



$k = 0$, $\phi = 1$	$k = 2\pi/a$, $\phi = 0$
$x_1 = \lambda$ (twice degenerated solution)	$\longrightarrow$ constant (flat band)
$x_3 = \lambda$	$\rightarrow -2\Delta + \sqrt{4\Delta^2 + \lambda^2}$
$x_5 = \lambda - 4\Delta$	$\rightarrow$
$x_2 = -\lambda$ (twice degenerated solution)	$\longrightarrow$ constant (flat band)
$x_6 = -\lambda$	$\rightarrow -2\Delta + \sqrt{4\Delta^2 + \lambda^2}$
$x_4 = -\lambda - 4\Delta$	$\rightarrow$

In Figure 8.15, we can see that when, as in the scenario (a), $\lambda < 2\Delta$, the two highest and lowest bands have the same width (2λ) and the overlap integral (λ) is small. This is quite a complicated function of distance (δ) between atoms, but in general it just results in a steady decrease, even though we should state that in this case, the atoms are so far apart they could almost be called independent. Indeed at the limit of this, the energies of the atoms resemble those of isolated atoms. In the second case denoted (b), where $\lambda > 2\Delta$, the two upper and lower bands have the same width (4Δ). The overlap integral is large and the distance (δ) between atoms is small.

8.4.3.4. *Study of band filling*

Let us look at how levels fill up with electrons in practical terms. The assumption is made that in directions other than the [100] shown in Figure 8.15, the bands are similarly distributed throughout the ranges of energies. This can be compared against other works, for example, G. Fishman where the curves are traced in the directions [100] and [111], and for the same scenario as in (a) $\lambda = \Delta$, and then when scenario (b) is looked at, Fishman uses $\lambda = 3\ \Delta$.

A priori, we should be able to place all $8N$ electrons into the system, as it was assumed that there were two electrons in each N cells, and that each atom liberates four electrons. As there are N cells, k takes on all N values with energy levels that have two places for electrons of opposing spin. Finally, the $8N$ electrons should fill, in both cases, the lowest $8N$ energy levels, as shown in Figure 8.15 for the four solutions x_4, x_6, and x_2 (twice degenerated). Each of these four solutions, in effect, contains N levels (equal to N values of k) filled with $2N$ electrons, such that the four solutions can take on all the $8N$ electrons spread over the whole extent of the bands.

The result is that in for case (a), as shown in Figure 8.15a, the medium is a conductor as the last two occupied levels, resulting from solution x_2, can easily mix their states with those of the empty levels from solution x_5. Conversely, in scenario (b), as shown in Figure 8.15b, the medium is an insulator (or semiconductor) as the last occupied levels (x_2) is far, in terms of energy, from $E_G = (2\lambda - 4\Delta)$ of the first empty level of x_5. This inhibits state mixing (that can only happen now with a high activation energy). In case (b) the two upper and lower bands have a width equal to 4Δ.

8.4.3.5. *Study of the wave functions: the s and p characters*

From the solution for the secular (otherwise termed characteristic) equation, x_i, it is a simple matter to obtain the coefficients denoted A_i and $A_i^{'}$ which give the linear combinations for the atomic orbitals that result in a determination of the wave function (Ψ) in the crystal.

The result, shown in Figure 8.16, is that the flat bands remain *p*-type (and are denoted *p*) over the whole of the k values (but with an evolution of the orbital linear combination $\varphi_{3p_{x,y,z}}$ and $\varphi'_{3p_{x,y,z}}$). When $k = 0$, the functions are strictly s or p; however, when $k \neq 0$ they are a mix of s and p, but remain mostly either *s*- or *p*-type, and are therefore respectively denoted s_0 or p_0.

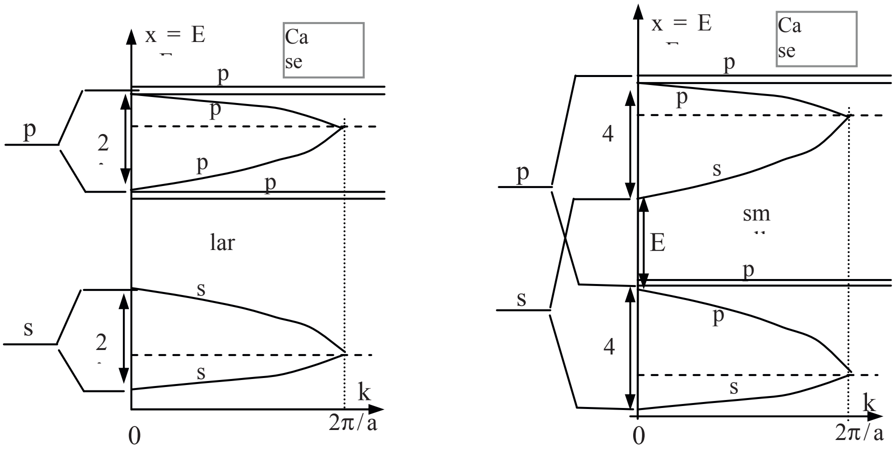


Figure 8.16. *Generation of wave functions from atomic orbitals for: (a) when δ is large (i.e. $\lambda < 2\Delta$); and (b) δ is small (i.e. $\lambda > 2\Delta$)*

8.4.3.6. *Representation of energy as a function of distances between atoms*

Figure 8.17 shows the energy diagram in direct space, i.e. in respect of the space between atoms.

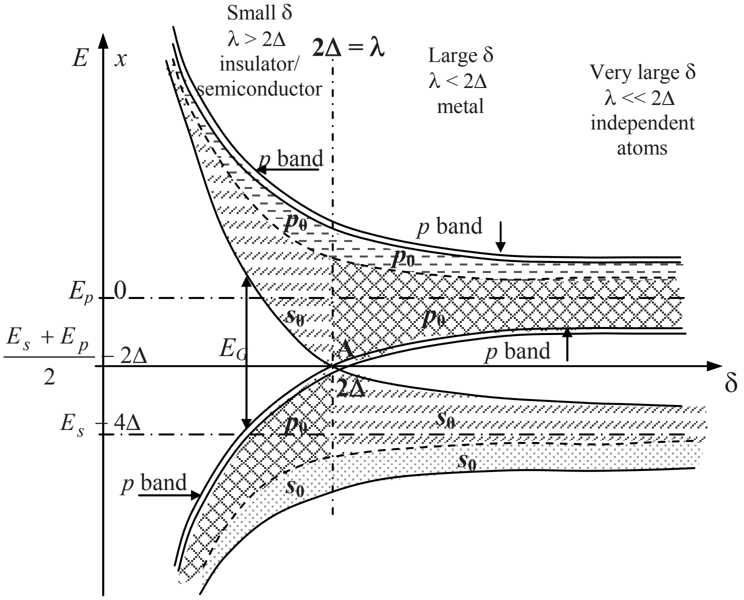


Figure 8.17. *Energy (E) (or the parameter $x = E - E_p$) and the bands as a function of distance (δ) between atoms*

The two places where $2\Delta = \lambda$, on either side of A, give us two types of behavior (metallic on the right where δ is large, and insulator on the left where δ is small). On the right there are two p_0 bands (definitely p when $k = 0$, and mostly p when $k \neq 0$) which are in the conduction band (CB), and two s_0 bands (definitely s when $k = 0$, mostly s when $k \neq 0$) which are in the VB. The high band for s_0 is in the CB. For the two flat bands (each doubly degenerated), the upper p band is always in the CB, while the lower band is in the CB on the right and in the VB on the left.

8.4.4. *Conclusion*

The results shown above come from using several simplifications, such as ignoring next-nearest neighbors and an imposed hybridization on the s- and p-states in the overlap integration. More elaborate theories have made it possible to get a more accurate vision of the system. The example opposite, drawn from Chelikowsky and Cohen, *Phys. Rev. B* 14, 556 (1976), shows the minimum of the CB and the maximum of the VB obtained or different k values, leading to an indirect gap.

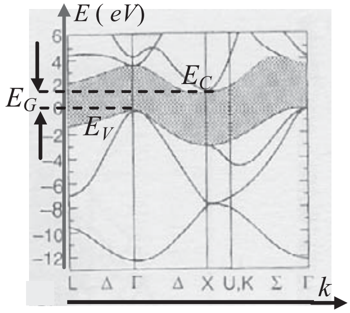


Figure 8.18. *Silicum band scheme in agreement with Chilikowsky and Cohen*

8.5. Problems

8.5.1. *Problem 1: strong bonds in a square 2D lattice*

For a square crystalline structure, which has a cell length denoted by a and a = 0.3 nm, there are identical atoms each tied to their neighbor by a single orbital (s in this case so as to simplify). In the approximation for the strong bond, the energy is always written such that $E = E_0 - \alpha - \beta \sum_m e^{-i\vec{k}\,\overrightarrow{a_m}}$, where $\overrightarrow{a_m}$ represents the vectors joining the reference atom to its nearest m neighbors (here $\beta > 0$, as we are dealing with s orbitals).

In reciprocal space, Γ, X and M denote the points respectively situated at the center of the zone (wave vector is zero), at the limit of the zone in the direction [10] (such that $\overrightarrow{\Gamma X} = \vec{G}_{10}/2$), and at the limit of the zone in the direction [11] (such that $\overrightarrow{\Gamma M} = \vec{G}_{11}/2$).

1) Determine the value of the energy at the points Γ, X and M.

2)

a) Give, for directions [10] and [11], the expressions for $E = f(k)$ resulting from a strong bond representation, and the expressions for $E^0 = f(k)$ for an approximation for free electrons.

b) In section 5.11, problem 2, there is: $\frac{\hbar^2}{2m}\frac{\pi^2}{a^2} \simeq 4.2$ eV (when $a \approx 0.3$ nm). For $\alpha = 1$ eV and $\beta = 0.5$ eV, give the numerical value that will be taken on by E_0 in order that the two curves $E = f(k)$ and $E^0 = f(k)$ go through the same origin with respect to Γ. Compare on the same scheme, for the directions [10] and [11], the evolution of the curves $E = f(k)$ and $E^0 = f(k)$.

3)

a) Show the form of the equi-energy curves around the points Γ.

b) Same as question 3*a* for the point *M*.

c) Give the specific equi-energy curve for $E = E_0 - \alpha$, and state to what the curve corresponds for a system of monovalent atoms.

4) Give the form of the effective mass around the point Γ.

5) Using the general expression for the state density function in 2D (see Chapter 2, problem 6), i.e. $Z(E) = \frac{1}{2\pi^2}\int_{L(E)} \frac{dL}{\left|\overrightarrow{\text{grad}_k E}\right|}$, give a qualitative indication of the form of the state density function of $Z(E)$.

Answers

1) For the square lattice, an atom situated at the origin of the lattice has four near neighbors giving rise to four vectors denoted $\overrightarrow{a_m}$ with the components:

$$\begin{cases} (\pm a, 0) \text{ along } x \\ (0, \pm a) \text{ along } y \end{cases}$$

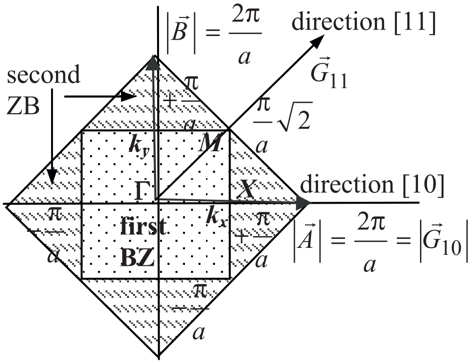


Figure 8.19. *Reciprocal lattice of the square structure*

The facing figure shows the reciprocal lattice obtained in problem 2 of Chapter 5 with the notation used therein.

By making k_x and k_y the components of the wave vector $\vec{k}$, the energy can be written as:

$$\mathrm{E} = \mathrm{E}_0 - \alpha - \beta \sum_m e^{-i\vec{k}\,\overrightarrow{a_m}} = \mathrm{E}_0 - \alpha - 2\beta\left[\cos k_x a + \cos k_y a\right]$$

For the point Γ, we have $\vec{k} = \vec{0}$, such that $k_x = k_y = 0$, from which $E(\Gamma) = E_0 - \alpha - 4\beta$. In terms of X, we therefore have $\vec{k} = \vec{k}_x = \overrightarrow{\Gamma X} = \vec{G}_{10}/2$, so that $k_x = \frac{\pi}{a}$ and $k_y = 0$, and hence $E(X) = E_0 - \alpha - 2\beta\left(\cos\pi + \cos 0\right) = E_0 - \alpha$.

For the point, we have $\vec{k} = \vec{k}_x + \vec{k}_y = \overrightarrow{\Gamma M} = \vec{G}_{11}/2$, such that $k_x = k_y = \frac{\pi}{a}$. The upshot is that $E(M) = E_0 - \alpha - 2\beta\left(\cos\pi + \cos\pi\right) = E_0 - \alpha + 4\beta$.

2)

a) In the direction [10] with respect to $\overrightarrow{\Gamma X}$, we have $k_y = 0$, and the energy along y is given by:

$-E_{[10]} = E_0 - \alpha - 2\beta\cos k_x a$ in the strong bond approximation;

$$-E^{0}_{[10]} = \frac{\hbar^2}{2m} k_x^2$$ in the free electron approximation.

In the direction [8.11] along $\overrightarrow{\Gamma M}$, we have $\vec{k} = \vec{k}_x + \vec{k}_y$ such that (square lattice) $k_x = k_y = \frac{|\vec{k}|}{\sqrt{2}}$ and the *y* energy is in the form:

$$-E_{[11]} = E_0 - \alpha - 4\beta \cos \frac{|\vec{k}|}{\sqrt{2}} a$$ for the strong bond approximation

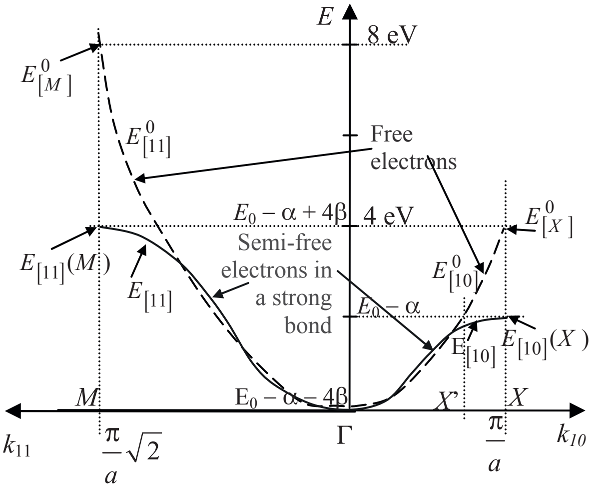
$$-E^{0}_{[11]} = \frac{\hbar^2}{2m}\left[k_x^2 + k_y^2\right] = \frac{\hbar^2}{2m} k^2$$ for the free electron approximation

b) Here we have $E(\Gamma) = E_0 - \alpha - 4\beta$ and with the free electrons, $E^{0}_{(\Gamma)} = \left[\frac{\hbar^2 k^2}{2m}\right]_{k=0} = 0$. So as to have the same origin for the energies at Γ, we can make $E(\Gamma) = 0$, and then $E_0 - \alpha = 4\beta$, so that with $\beta = 0.5$ eV, we find $E_0 - \alpha = 2$ eV. When $\alpha = 1$ eV, we now have $E_0 = 3$ eV. Numerically, with $\frac{\hbar^2}{2m}\frac{\pi^2}{a^2} \simeq 4.2$ eV, we have:

– $E_{[10]}(X) = E(X) = E_0 - \alpha = 2$ eV and $E^{0}_{[X]} = \frac{\hbar^2}{2m} k_x^2 = \frac{\hbar^2}{2m}\frac{\pi^2}{a^2} = 4.2$ eV

– $E_{[11]}(M) = E(M) = E_0 - \alpha + 4\beta = 4$ eV and $E^{0}_{[M]} = \frac{\hbar^2}{2m}[k_x^2 + k_y^2] = \frac{\hbar^2}{m}\frac{\pi^2}{a^2} = 8.4$ eV

In the graphical representation below, we can see that around the point Γ the two curves are in good agreement. However, at the limits of the zone (points *X* and *M*), the free electrons are no longer influenced by bonds with the lattice. The band width is equal to $\Delta E = E(M) - E(\Gamma) = 8\beta$. This scheme should also be compared with that shown in Chapter 5, problem 2 for semi-free electrons in a weak bond.



3) The equi-energy curves are defined by the relation:

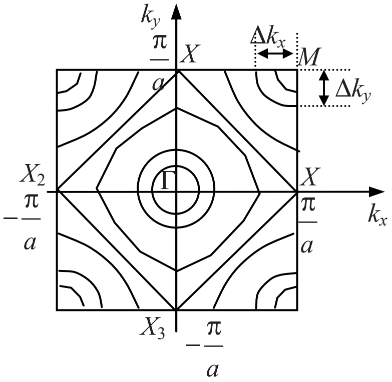
$$E = E_0 - \alpha - 2\beta\left[\cos k_x a + \cos k_y a\right] = \text{constant} \qquad [8.16]$$

a) In the neighborhood of Γ where $k_x \simeq k_y \simeq \varepsilon \to 0$,, we can write that $\cos k_x a \simeq 1 - \frac{(k_x a)^2}{2}$ and $\cos k_y a \simeq 1 - \frac{(k_y a)^2}{2}$. Placing this into the preceding equation [8.16] gives $\frac{E - E_0 + \alpha + 4\beta}{\beta a^2} = k_x^2 + k_y^2$. In the reciprocal k space, E = constant, which implies that $k^2 = k_x^2 + k_y^2$ = constant. In the neighborhood of Γ, the equi-energy curves are circles centered about Γ.

b) In the neighborhood of M (where $\overrightarrow{\Gamma M} = \vec{k}_{\Gamma M}$ and has the components $k_x = \frac{\pi}{a}$ and $k_y = \frac{\pi}{a}$) the vector $\vec{k}$ has components k_x and k_y which can be written such that $k_x = \frac{\pi}{a} - \Delta k_x$ and $k_y = \frac{\pi}{a} - \Delta k_y$, where $\Delta k_x \equiv \Delta k_y \simeq \varepsilon \to 0$. From this, $\cos k_x a = \cos(\pi - a\Delta k_x) = -\cos a\Delta k_x \simeq \frac{(a\Delta k_x)^2}{2} - 1$, and similarly, $\cos k_y a \simeq \frac{(a\Delta k_y)^2}{2} - 1$. Equation [8.16] thus gives:

$$\Delta k_x^2 + \Delta k_y^2 = \frac{E_0 + 4\beta - \alpha - E}{\beta a^2} = \text{constant}$$

when E = constant. The equi-energy lines in reciprocal space are circles centered about $k_x = \frac{\pi}{a}$ and $k_y = \frac{\pi}{a}$ (components of $\overrightarrow{\Gamma M}$) and have a radius $\sqrt{\Delta k_x^2 + \Delta k_y^2} = \frac{\sqrt{E_0 + 4\beta - \alpha - E}}{a\sqrt{\beta}}$.



c) When $E = E_0 - \alpha$, equation [8.16] gives $\cos k_x a = -\cos k_y a$. As $\cos\alpha = -\cos\beta$ implies that $\beta = \pm\pi \pm \alpha$, we should here have $k_y = \pm k_x \pm \frac{\pi}{a}$. In the representation of $k_y = f(k_x)$ the equation has straight slopes at ±1 and an ordinate $\pm\frac{\pi}{a}$ at the origin. The equi-energy curve, circular around Γ, transforms (while passing through the intermediate forms as shown in the preceding scheme) with E tending to $(E_0 - \alpha)$ to finally bring the lines together. This results in the first zone of the square having high points at X, X_1 ($k_x = 0$, $k_y = \frac{\pi}{a}$), X_2 ($k_x = -\frac{\pi}{a}$, $k_y = 0$), and X_3 ($k_x = 0$, $k_y = -\frac{\pi}{a}$). This result is quite normal given that we have obtained above $E_{[10]}(X) = E(X) = E_0 - \alpha$. The side of the square, equal to the half-length of the diagonal of a square with side $\frac{\pi}{a}$ represents the Brillouin zone and thus is equal to: $\frac{\pi}{a}\sqrt{2}$. The surface of this square is thus equal to the half-surface of the Brillouin zone and can only contain $N/2$ cells when the Brillouin zone contains N. This square surface of side $\frac{\pi}{a}\sqrt{2}$ can therefore contain N electrons, and is such that the sides constitute the limit of the Fermi curve if the atoms making up the lattice are monovalent.

4) Around the point Γ where k (and therefore k_x and k_y) is small, we can perform a limited development of the cosine functions involved in the expression for the energy:

$$E = E_0 - \alpha - 2\beta\left[\cos k_x a + \cos k_y a\right]$$

$$\simeq E_0 - \alpha - 2\beta\left[2 - \frac{k_x^2 a^2}{2} - \frac{k_y^2 a^2}{2}\right] = E_0 - \alpha - \beta\left[4 - k^2 a^2\right];$$

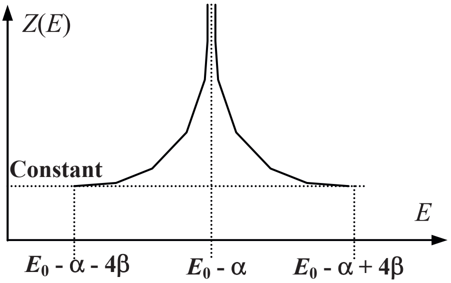
and from this it can be deduced that $\frac{\partial^2 E}{\partial k^2} = 2\beta a^2$, such that $m^* = \frac{\hbar^2}{2\beta a^2}$.

5) In the neighborhood of point Γ (energy $E = E_0 - \alpha - 4\beta$), the dispersion curve is close to that of the free electrons and $Z(E)$ should be pretty much constant in this region (as $Z(E)$ is a constant in 2D with respect to free electrons, as detailed in Chapter 2, problems 3 and 6).

In the neighborhood of E = (E_0 - α), we can use the relationship $Z(E) = \frac{1}{2\pi^2}\int_{L(E)} \frac{dL}{\left|\overrightarrow{\text{grad}}_k E\right|}$ where $L(E)$ is the equi-energy curve.

For free electrons, the equi-energy curve is given by $L'(E) = 2\pi\ \Gamma X'$ (circular equi-energy curve with radius Γ*X*', where *X*' is defined in the scheme in problem 2 and is such that Γ*X*' < Γ*X*). For electrons in strong bonds, the equi-energy curve has a length $L(E) = 4\ \Gamma X\sqrt{2}$ (square of the equi-energy curve, the side of the square being equal to $\Gamma X\sqrt{2}$). As Γ*X* > Γ*X*', in principle, we have $L(E) > L'(E)$. All is at least of the same order of size, and the exact calculation is not of interest given the importance in the variation of $|\overrightarrow{\text{grad}}_k E|$ between the two approximations, as we shall see.

In the neighborhood of $E = (E_0 - \alpha)$, the slope of the curve $E = f(k)$ is considerable for free electrons, but is practically zero for (tangent horizontal in *X*) for electrons in a strong bond so that in this case $\int_{L(E)} \frac{dL}{\left|\overrightarrow{\text{grad}}_k E\right|}$ is considerably greater than for cases with free electrons. $Z(E)$ for these electrons in a strong bond increases in size (at a point of singularity).



Beyond $E = (E_0 - \alpha)$, and on coming closer to the point M (energy $E = E_0 - \alpha + 4\beta$), the equi-energy curves for electrons in strong bonds tends to that of free electrons (circular arcs in the corners of the first Brillouin zone), and it can be expected that the state density functions for the two representations will tend towards one another. This ends up giving us the representation shown on the right.

8.5.2. *Problem 2: strong bonds in a cubic centered or face centered lattices*

This question first deals with a cubic centered (cc) crystalline structure (also called case 1) and then a face centered cubic (fcc) structure (called case 2). The atoms in both cases are all identical and, so as to make the problem simpler, tied to neighboring atoms only by *s*-orbitals. In the strong bond approximation, the energy is still written using $E = E_0 - \alpha - \beta \sum_m e^{-i\vec{k}\,\overrightarrow{a_m}}$, where $\overrightarrow{a_m}$ represents the vectors joining the reference atom (at the origin) to its nearest m neighbors (and here, $\beta > 0$ as we are dealing in *s*-orbitals). We can then go on to use the results from Chapter 6 on the cc structure (see section 6.6.1, problem 1) and on the fcc structure (sections 6.2 and 6.3).

1a and 2a) Establish the dispersion relation for both cases.

1b and 2b) Determine, for both cases, the energy values at the center of the zone, i.e. $E(\Gamma)$, at the point X at the intersection between direction [100] and the first Brillouin zone, i.e. $E(X)$, at the point M at the intersection between direction [110] and the first Brillouin zone, i.e. $E(M)$, and at the point W at the intersection between direction [111] and the first Brillouin zone, i.e. $E(W)$.

1c and 2c) Indicate for both cases the width of the band ΔE and establish the expression for m^* around the point Γ.

Answers

1)

a) For the cc structure, the reference atom (at the origin) at the highest point of the cube has eight near neighbors situated at the center of the eight adjacent cells. The components of the vectors $\vec{a}_m$ joining the origin-based atom with its closest neighbors is thus of the form (associating two by two and including components with opposing signs):

$$\vec{a}_m \begin{cases} d/2 & -d/2 & -d/2 & d/2 & d/2 & -d/2 & d/2 & -d/2 \\ d/2 & -d/2 & d/2 & -d/2 & -d/2 & d/2 & d/2 & -d/2 \\ d/2 & -d/2 & d/2 & -d/2 & d/2 & -d/2 & -d/2 & d/2 \end{cases}$$

Under these conditions the energy is written (with eight terms) as:

$$\begin{aligned} E = E_0 - \alpha - \beta \Big\{ & e^{-i\left[d/2\left(k_x + k_y + k_z\right)\right]} + e^{i\left[d/2\left(k_x + k_y + k_z\right)\right]} \\ & + e^{-i\left[d/2\left(-k_x + k_y + k_z\right)\right]} + e^{i\left[d/2\left(-k_x + k_y + k_z\right)\right]} \\ & + e^{-i\left[d/2\left(k_x - k_y + k_z\right)\right]} + e^{i\left[d/2\left(k_x - k_y + k_z\right)\right]} \\ & + e^{-i\left[d/2\left(k_x + k_y - k_z\right)\right]} + e^{i\left[d/2\left(k_x + k_y - k_z\right)\right]} \Big\} \end{aligned}$$

from which it is deduced that:

$$\begin{aligned} E = E_0 - \alpha - 2\beta \Big\{ & \cos \frac{d\left(k_x + k_y + k_z\right)}{2} + \cos \frac{d\left(-k_x + k_y + k_z\right)}{2} \\ & + \cos \frac{d\left(k_x - k_y + k_z\right)}{2} + \cos \frac{d\left(k_x + k_y - k_z\right)}{2} \Big\} \end{aligned}$$

b) At the point Γ, we have $k_x = k_y = k_z = 0$, and $E(\Gamma) = E_0 - \alpha - 8\beta$. The point X (≡ B on the figure shown in problem 4 of section 6.6.1) is such that $\overrightarrow{\Gamma X}$ ($\equiv \overrightarrow{OB}$ in the figure of problem 4, section 6.6.1) has the components given by: $k_x = \frac{2\pi}{d}$, $k_y = 0$, $k_z = 0$. The result is that $E(X) = E_0 - \alpha + 8\beta$.

The point M is such that $\overrightarrow{\Gamma M}$ ($\equiv \overrightarrow{OG}/2$ in the figure of section 6.6.1) has the components $k_x = \frac{\pi}{d}$, $k_y = \frac{\pi}{d}$, $k_z = 0$. The upshot from all this is that:

$$E(M) = E_0 - \alpha - 2\beta[\cos\pi + \cos 0 + \cos 0 + \cos\pi] = E_0 - \alpha.$$

The point W is such that $\overrightarrow{\Gamma W}$ ($\equiv \overrightarrow{OF}$ from the same figure) has the components $k_x = \frac{\pi}{d}$, $k_y = \frac{\pi}{d}$, $k_z = \frac{\pi}{d}$. The result is that $E(W) = E_0 - \alpha$.

c) Here we have $\Delta E = E(X) - E(\Gamma) = E(M) - E(\Gamma) = 16\,\beta$. Around Γ where k is small, we can use the limited development of $\cos\theta \simeq 1 - \frac{\theta^2}{2}$ which makes it possible to write the energy in the neighborhood of Γ as:

$$E \simeq E_0 - \alpha - 2\beta\left[4 - \frac{1}{2}\frac{d^2}{4}4k^2\right] = E_0 - \alpha - 8\beta + \beta d^2 k^2.$$

Thus, near Γ, the effective mass, which is such that $m^* = \frac{\hbar^2}{\frac{\partial^2 E}{\partial k^2}}$ and here $\frac{\partial^2 E}{\partial k^2} = 2\beta d^2$, is therefore give by $m^* = \frac{\hbar^2}{2\beta d^2}$.

NOTE ON THE SUBJECT OF QUESTION 1A.– The sought-after energy of problem 1a can be written with the following notation:

$$E = E_0 - \alpha - \beta \sum_{\varepsilon_x = \pm, \varepsilon_y = \pm, \varepsilon_z = \pm} e^{i d/2(\varepsilon_x k_x + \varepsilon_y k_y + \varepsilon_z k_z)}.$$

And, in general terms, it is possible to state that:

$$\begin{aligned}8\cos a\cdot\cos b\cdot\cos c &= 8\frac{e^{ia}+e^{-ia}}{2}\frac{e^{ib}+e^{-ib}}{2}\frac{e^{ic}+e^{-ic}}{2}\\ &=(e^{ia}+e^{-ia})(e^{ib}+e^{-ib})(e^{ic}+e^{-ic})\\ &=\sum_{\varepsilon_a=\pm,\varepsilon_b=\pm,\varepsilon_c=\pm} e^{i(\varepsilon_a a+\varepsilon_b b+\varepsilon_c c)},\end{aligned}$$

The energy for question 1a can also be written in the form:

$$E=E_0-\alpha-8\beta\cos\frac{k_x d}{2}\cos\frac{k_y d}{2}\cos\frac{k_z d}{2}.$$

The results from problems 1b and 1c remain unchanged.

2)

a) For the fcc structure, the reference atom (origin) at the summit of a cube has closest neighbors at the centers of the three faces of eight adjacent cells. The three faces are along the planes *Oxy*, *Oxz* and *Oyz*. Each node at the center of the faces is shared with two cells and this gives in all $\frac{3\times 8}{2}=12$ nodes that are immediate neighbors. We can equally state that in each plane, there are four nodal "centers" so that in all there are 4 × 3 = 12 nodes that are immediate neighbors. The components of the $\vec{a}_m$ vectors that join the origin-based node to the 12 neighbors are given by (again, grouped two by two):

$$\vec{a}_m\left\{\begin{array}{cc|cc|cc|cc|cc|cc} d/2 & -d/2 & d/2 & -d/2 & d/2 & -d/2 & d/2 & -d/2 & 0 & 0 & 0 & 0\\ d/2 & -d/2 & -d/2 & d/2 & 0 & 0 & 0 & 0 & d/2 & -d/2 & d/2 & -d/2\\ 0 & 0 & 0 & 0 & d/2 & -d/2 & -d/2 & d/2 & d/2 & -d/2 & -d/2 & d/2\end{array}\right.$$

The energy written under these conditions (with 12 terms) is:

$$\begin{aligned} E = E_0 - \alpha - \beta \Big\{ & e^{i\left[d/2\left(k_x + k_y\right)\right]} + e^{-i\left[d/2\left(k_x + k_y\right)\right]} \\ & + e^{i\left[d/2\left(k_x - k_y\right)\right]} + e^{-i\left[d/2\left(k_x - k_y\right)\right]} \\ & + e^{i\left[d/2\left(k_x + k_z\right)\right]} + e^{-i\left[d/2\left(k_x + k_z\right)\right]} \\ & + e^{i\left[d/2\left(k_x - k_z\right)\right]} + e^{-i\left[d/2\left(k_x - k_z\right)\right]} \\ & + e^{i\left[d/2\left(k_y + k_z\right)\right]} + e^{-i\left[d/2\left(k_y + k_z\right)\right]} \\ & + e^{i\left[d/2\left(k_y - k_z\right)\right]} + e^{-i\left[d/2\left(k_y - k_z\right)\right]} \Big\} \end{aligned}$$

$$\begin{aligned} E = E_0 - \alpha - 2\beta \Big\{ & \cos\left[\frac{d}{2}\left(k_x + k_y\right)\right] + \cos\left[\frac{d}{2}\left(k_x + k_z\right)\right] \\ & + \cos\left[\frac{d}{2}\left(k_y + k_z\right)\right] + \cos\left[\frac{d}{2}\left(k_x - k_y\right)\right] \\ & + \cos\left[\frac{d}{2}\left(k_x - k_z\right)\right] + \cos\left[\frac{d}{2}\left(k_y - k_z\right)\right] \Big\} \end{aligned}$$

b) At the point Γ, we again have $k_x = k_y = k_z = 0$ and $E(\Gamma) = E_0 - \alpha - 12\beta$. The point X ($\equiv H'$ in Figure 6.4) is such that $\overrightarrow{\Gamma X}$ ($\equiv \overrightarrow{OH'}$ in Figure 6.4) has the components $k_x = \frac{2\pi}{d}$, $k_y = 0$, $k_z = 0$. The result is that:

$$E(X) = E_0 - \alpha - 2\beta\ (\cos\pi + \cos\pi + \cos 2\pi + \cos\pi + \cos\pi + \cos 2\pi) = E_0 - \alpha + 4\beta.$$

The point M is such that $\overrightarrow{\Gamma M}$ ($\equiv \overrightarrow{OM}$ Figure 6.4) and has the components $k_x = \frac{3\pi}{2d}$, $k_y = \frac{3\pi}{2d}$, $k_z = 0$. The result is that:

$$E(M) = E_0 - \alpha - 2\beta\left(\cos\frac{3\pi}{2} + \cos 2\pi + 4\cos\frac{3\pi}{4}\right) = E_0 - \alpha + 2\beta\left(2\sqrt{2} - 1\right).$$

The point W is such that $\overrightarrow{\Gamma W}$ $(\equiv \overrightarrow{OG}'$, Figure 6.4) and has the components $k_x = \frac{\pi}{d}$, $k_y = \frac{\pi}{d}$, $k_z = \frac{\pi}{d}$. The result is that

$$E(W) = E_0 - \alpha - 2\beta\ \left[3\cos\pi + 3\cos 2\pi\right] = E_0 - \alpha$$

c) For the fcc system, we have $\Delta E = E(X) - E(\Gamma) = 16\beta$ (which does not accord with $\Delta E = 2Z\beta$ because in this case where $Z = 12$ we now have $\Delta E = 24\beta$).

The use of the limited cosine development in the neighborhood of Γ gives:

$$\begin{aligned} \mathrm{E} \ \simeq\ \mathrm{E}_0 - \alpha - 2\beta\Bigg\{6 - \frac{1}{2}\frac{d^2}{4}\left[\left(k_x + k_y\right)^2 + \left(k_x + k_z\right)^2\left(k_y + k_z\right)^2\right] \\ - \frac{1}{2}\frac{d^2}{4}\left[\left(k_x - k_y\right)^2 + \left(k_x - k_z\right)^2\left(k_y - k_z\right)^2\right]\Bigg\}. \end{aligned}$$

From this we deduce that

$$E(\Gamma) \simeq E_0 - \alpha - 12\beta\ -2\beta\left\{-\frac{1}{2}\frac{d^2}{4}4k^2\right\} = E_0 - \alpha - 12\beta + \beta k^2 d^2.$$

Once again, we obtain the effective mass: $m^* = \dfrac{\hbar^2}{2\beta d^2}$.

NOTE ON THE SUBJECT OF PROBLEM 2A.– In a manner similar to that of problem 1a, we can note that for problem 2a, the following relation can be used:

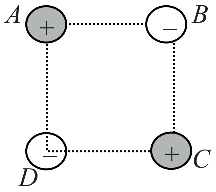
$$\begin{aligned} 4\cos a \cdot \cos b &= 4\frac{e^{ia} + e^{-ia}}{2}\frac{e^{ib} + e^{-ib}}{2} \\ &= (e^{ia} + e^{-ia})\,(e^{ib} + e^{-ib}) \\ &= \sum_{\varepsilon_a = \pm, \varepsilon_b = \pm} e^{i\left(\varepsilon_a a + \varepsilon_b b\right)}. \end{aligned}$$

This makes it possible to state E for the fcc system:

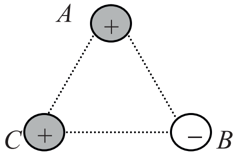
$$E = E_0 - \alpha - 4\beta\left(\cos\frac{k_x d}{2}\cos\frac{k_y d}{2} + \cos\frac{k_x d}{2}\cos\frac{k_z d}{2} + \cos\frac{k_y d}{2}\cos\frac{k_z d}{2}\right).$$

The results for questions 2b and 2c are identical.

Comment on alternating and non-alternating structures



Alternate structure
All closed pathways (loops) joining nearest neighbors contain an even number of atoms (square lattice, simple or cc for example). If the signs of the wave functions between A and B are reversed, and the between B and C, then C and D, the sign between the nearest neighbors D and A remains alternate. In such a structure, the most anti-bonding bond is that which has an energy equal to $E_0 - \alpha + Z\beta$, where Z is the number of nearest neighbors.



Non-alternate structure
There are pathways which are based on an odd number of atoms, for example the triangular pathway in the [111] plane of a fcc lattice. In this system the alternation of the wave function signs on a route from between A and B, then B and C, means that the sign between C and A is non-alternate. In such a structure, the most anti-bonding state thus has an energy which is below $E_0 - \alpha + Z\beta$ (and we find $E(X) = E_0 - \alpha + 4\beta$ in place of $E(X) = E_0 - \alpha + 12\beta$). The width of the band B nevertheless remains proportional to β. In the fcc structure, $\Delta E \equiv B = 16\ \beta$. As $m^* = \frac{\hbar^2}{2\beta d^2}$, the result is that $m^* \propto \frac{1}{B}$ and the mobility $\mu \propto B$ (with the same law of variation for m^* and μ as a function of B as those established in section 8.1.3 for alternate systems; only the coefficients are different).

NOTE.– The structure of carbon diamond (based on two fcc imbricated lattices), along with its tetrahedral bonds and coordination number equal to 4, behaves different to the fcc structure.

Chapter 9

Limits to Classical Band Theory: Amorphous Media

This chapter will evaluate the limits of classical band theory, for which we have, up until now:

– assumed that there were no electronic defaults such as dangling bonds which normally arise in column IV semiconductors at the edges of finite volumes, where the covalent bond arising from one atom (C, Si) dangles away from the bulk of the material;

– neglected electron–electron (Fermi gas, or free electron gas) interactions; and

– only looked at perfect, crystalline solids.

9.1. Evolution of the band scheme due to structural defects (vacancies, dangling bonds and chain ends) and localized bands

Figure 9.1 takes up where the band scheme for tetrahedral carbon shown in Figure 8.10b left off and gives a summary of the formation of valence and conduction bands in a perfectly ordered medium. As detailed in Chapter 8, the initial s^2p^2 configuration gives rise to the formation of four molecular sp^3 orbitals, each one yielding bonding (σ) and anti-bonding (σ^*) orbitals. On moving to the solid state, there is a breakdown in the σ and σ^* levels into bands, with the former generating the valence band, and the latter the conduction band.

Nevertheless, in a real crystal, its finite dimensions result in bond defaults. At the end of a chain at the edge of the crystal, a carbon atom is only tied to three other carbons, and this leaves an unsatisfied sp^3 bond. This is called a dangling bond, and intrinsically contains one electron, but is electrically neutral. The electron is therefore on the E_{sp3} level, and the dangling bond has a localized (E_{Loc}) level associated with it that is situated in the middle of the gap, taking into account the construction of the permitted bands (Figure 9.1).

There are other structural faults that can also give rise to similar levels in a real crystal. These include site vacancies (generated during the preparation of the crystal), dangling bonds introduced through a physical treatment (irradiation; ionic implantation which results in bond rupture as the ions pass through the material).

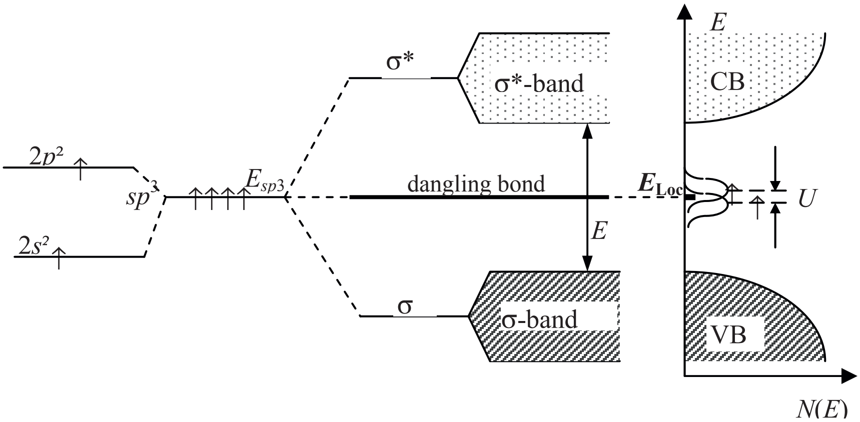


Figure 9.1. *The formation of localized levels associated with dangling bonds in tetrahedral carbon*

The presence of more than one structural default from dangling bonds results in an opening (for example due to the disorder created by fluctuations in bond angles) in levels in the middle of the gap as a default band. The precise position of these bands is in effect tied to the relaxation phenomena produced on formation of the defaults. It is also related to the exact origin (valence of conduction band) of the states giving rise to the defaults (dangling bond).

The lower band at the middle of the gap in Figure 9.1 thus corresponds to a neutral dangling bond that contains a single electron. It is a donor-type band and is thus neutral in the occupied state. The upper band at the middle of the gap corresponds to the same default but exhibits a different charge state (that is that it has received an extra electron). This is an acceptor-type band that is neutral when empty. The energy difference between the two types of default is equal to the Hubbard correlation energy (U) which takes on a value given by $U = \langle q^2/4\pi\varepsilon_0\varepsilon_r r_{12}\rangle$. r_{12} Denotes the distance between two electrons on the same site. An average is made over all possible configurations. We shall now try to determine the general effect of such electronic repulsion.

9.2. Hubbard bands and electronic repulsions. Mott metal–insulator transition

9.2.1. *Introduction*

In the discussion up to now on band theory in this book, it has always been assumed that each electron moved in an average potential resulting from atomic nuclei and other electrons considered together, rather than including individual electron–electron interactions. This assumption can seem erroneous, especially when realizing that the electrostatic repulsive energy (Coulomb) is of the same order as the kinetic energy of electrons in the Fermi gas. It is thus surprising that band theories that just account for global interactions of electrons with the lattice can have the success that they have had. There are possible reasons for this, for example the assumption that the electrostatic interactions between two electrons can be screened by a conjugation of moving electrons that form a potential sphere. Other examples include the Pauli exclusion principle that largely decreases the probability of proximity between electrons. It can equally be thought that independent particles are not electrons but *quasi*-particles that are tied to their complementary particles (holes).

Nevertheless, there are cases where the default method is to take the independent interactions between electrons into account. Using this method it is possible to see the transition between insulating and metallic states simply by modifying the "electronic concentration". It is also possible to see a state predicted by band theory to be metallic to in fact be an insulator. This effect can be seen as a result of the effect of competition between kinetic and potential electron energies. For example, the metallic state is favored if the gain in kinetic energy is equal to moving electrons from the electronic state to the available empty atomic Bloch states in the conduction band (and as a consequence close to the Fermi level E_F). Indeed, Heisenberg's principle also explains this behavior because if x increases with delocalization, the V decreases; the cost in potential energy is equal to that given to the electrons so that they can overcome the electrostatic potential energy which concentrates them in the lattice. It should be noted also that this occurs with an electrostatic potential energy that is all the greater as the permittivity of the medium decreases (hence leading to an insulating material). Conversely, if the material tends towards the metallic state ($\varepsilon \rightarrow \infty$), then the potential energy goes towards zero. This condition will be used in section 9.2.3 to derive the criteria for the Mott transition.

NOTE.– The term Fermi gas is reserved for media in which we neglect electron–electron interactions, while the term Landau liquid is used in situations where these interactions are accounted for.

9.2.2. *Model*

Using the Hubbard model and hypotheses, we assume that the only electron–electron interactions that are of interest are those due to two electrons on the same site (the same atom when looking at the example of a chain of atoms). This repulsive energy (Hubbard energy) can be evaluated. It can then be shown that it is considerable in certain situations, and makes possible an interpretation of the origin of some metal–insulator transitions. In order to carry out this evaluation, we shall use the example of a line of alkaline metals.

In band theory, until now, we have considered that each electron existed in an average potential resulting from a collection of atoms and other electrons. In the case of alkali metals (Li, Na, K,…), which have one free electron per atom, the transfer of an electron from one atom to its neighbor through a conduction band occurs via electronic levels situated just above the Fermi level (E_F) and the energy utilized is extremely small, of the order of a fraction of a meV.

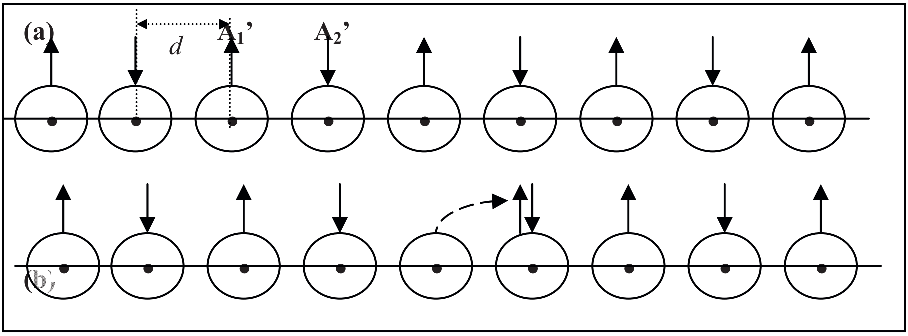


Figure 9.2. *Highlighting electronic repulsions in a chain of atoms with s-orbitals*

The alkali metals shown in Figure 9.2a can be used to evaluate the problem. We will assume that overlapping between atoms is poor and the transport of electrons from one atom to the next requires a great deal of energy. Movement of an electron thus generates a supplementary repulsive energy which can be estimated by:

– calculating the ionization energy required (I_p) to separate an electron from the atom A_1' to which it is attached which subsequently becomes A_1 (this change is shown in going from Figure 9.2a to Figure 9.2b);

– and calculating the energy recovered, or the electron affinity (χ) when the free electron is placed on the independent, adjacent atom A_2', which subsequently becomes A_2.

The total energy thus required, equivalent to the repulsive energy, is $U_H = I_p - \chi$. For hydrogen $I_p = 13.6$ eV and $\chi = 0.8$ eV, and thus $U_H = 12.8$ eV, showing how U_H can attain a relatively high value of several eV.

Mott showed how the repulsive energy can be calculated using r_{12}, which represents the distance between two electrons on the same site or atom, and $\psi(r)$ which is the wave function corresponding to the value proposed at the end of section 9.1:

$$U_H = \iint \frac{e^2}{4\pi\varepsilon_0 r_{12}} \left|\psi(r_1)\right|^2 \left|\psi(r_2\right|^2 dr_1 dr_2 .$$

Practically speaking, this energy is particularly important with respect to transition metal oxides, such as NiO, for which electron transport occurs via *d*-orbitals and can be written as:

$$\mathrm{Ni}^{2+} + \mathrm{Ni}^{2+} \rightarrow \mathrm{Ni}^{3+} + \mathrm{Ni}^{+}$$

For a chain of alkali metals, however, the same electron transfer, via *s*-orbitals, is written:

$$\mathrm{Cs} + \mathrm{Cs} \rightarrow \mathrm{Cs}^{+} + \mathrm{Cs}^{-}$$

In Figure 9.2, $A_1' \equiv A_2' \equiv \mathrm{Cs}$ while $A_1 \equiv \mathrm{Cs}^+$ and $A_2 \equiv \mathrm{Cs}^-$. When the arrows are removed from Figure 9.2 that represent the division of electrons throughout a chain of atoms, we can consider that for NiO,

$$A_1' \equiv A_2' \equiv \mathrm{Ni}^{2+},\quad A_1 \equiv \mathrm{Ni}^{3+} \text{ and } A_2 \equiv \mathrm{Ni}^{+}$$

Placing an electron on a Ni^{2+}, to form a Ni^+ ion, would require the energy given by $U_H = I - \chi$ if the Ni^+ and Ni^{3+} ions, at positions A_2 and A_1 in Figure 9.2b, respectively, are sufficiently far apart, so that it can be assumed that the transported electron passes through a free state, and so that its energy E_n at the level $n \rightarrow \infty$ tends towards 0, as do the successive energies I_p and $-\chi$, as described above.

Energy levels of isolated ions can be represented in terms of $-I_p$ (the energy of an orbital which loses an electron, i.e. Ni^{3+} or A_1 in Figure 9.2) and $-\chi$ (the energy of a supplementary electron situated on Ni^+ or A_2 in Figure 9.2). When the ions are well separated, as shown in the far left part of Figure 9.3, each energy level is separated by $U_H = I_p - \chi$ which appears as a band gap between the upper and lower levels, the former having received an electron, the latter having lost one.

On bringing the ions closer to one another, as described in going from the left to right side of Figure 9.3, transport by charge carriers becomes possible via the permitted bands which start to form. These newly formed discrete bands give rise to permitted bands (the Hubbard bands), upper level bands of electrons (in which Ni^+ or A_2 can be found) and lower level bands containing holes (where Ni^{3+} or A_1 of Figure 9.2b can be found).

As the size (B) of the bands grows with increasing proximity of atoms, the difference $U_H - B$ decreases and eventually disappears when B reaches U_H (Figure 9.3). Beyond this value – obtained when the atoms are close enough to each other – the upper and lower Hubbard bands overlap and the band gap is removed; this point is also known as the Mott–Hubbard transition from an insulator to metallic state.

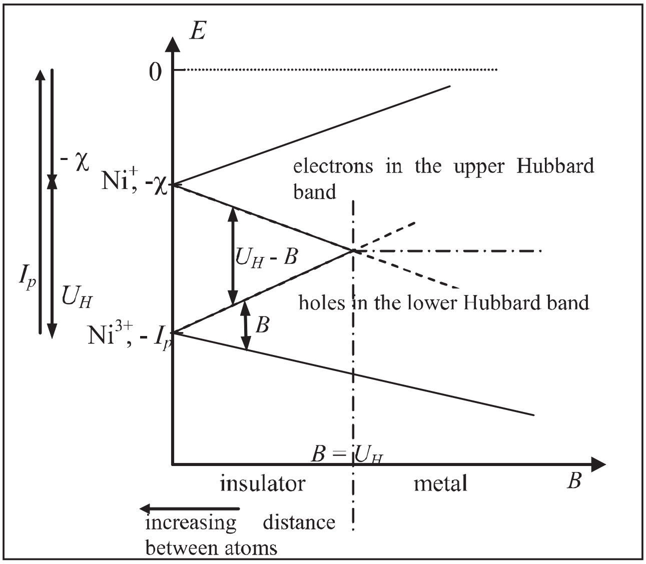


Figure 9.3. *Evolution of the Hubbard bands as a function of band size (B): B = 0 for atoms far apart; B = U_H, where the band gap $U_H - B$ disappears to give a metal–insulator transition*

9.2.3. *The Mott metal–insulator transition: estimation of transition criteria*

Different theories have been elaborated to establish, in a quantitative manner, the parameters surrounding transitions from insulator to metallic states (see problem 9.5.1). As we have seen in section 9.2.1, this transition can result from competition between localization effects, themselves resulting from electron Fermi kinetic energies (E_F) and the electrostatic energies to which the electrons are subject.

To take into account the effect of the medium in which the electrons reside, we introduce into the expression for electrostatic potential a distance of action written as K_1a^*. K_1 is a constant that takes into account the imprecision that reigns over much of the knowledge concerning interaction distances (and can be thought of as a rule of thumb), and is dependent on the hypotheses made on the various phenomena that can act, such as polarization (where $K_1 = \varepsilon_r$) or screening as detailed in section 9.5.1. a^* is the distance that should have a form similar to a_1, in that $a_1 = [\varepsilon_0 h^2]/[\pi m e^2]$), and is the radius of the first Bohr orbital (with the condition of taking into account the interaction with the lattice which transforms the mass of an electron to an effective mass, i.e. m^*. This transforms the equation to $a^* = [\varepsilon_0 h^2]/[\pi m^* e^2]$, so that $a^* = a_1(m/m^*)$.

The electrostatic potential energy (which tends to localize the electrons) must therefore be written as $W_{el} = \frac{e^2}{4\pi\varepsilon_0 K_1 a^*}$.

For its part, the kinetic energy of metallic electrons placed on the Fermi level (E_F) is written as $E_{cF} = \frac{\hbar^2}{2m^*}k_F^2$ (where k_F is the wave number for the Fermi level and where interactions between the electrons and the lattice are simply taken to account through m^*). The number of electrons (N) contained in the Fermi sphere is equal to the number of elementary cells that are present multiplied by two so as to account for spin. This gives us:

$$N = 2\frac{(4/3)\pi k_F^3}{(2\pi/L)^3} = \frac{L^3}{3\pi^2}k_F^3.$$

By introducing the electronic concentration, $N_e = N/L^3$, we have $k_F = (3\pi^2 N_e)^{1/3}$ and then substituting into E_{cF}:

$$E_{cF} = \frac{\hbar^2}{2m^*}(3\pi^2 N_e)^{2/3}.$$

The gain in kinetic energy by the electron gas is thus $E_{cF} = K_2 \frac{h^2 N_e^{2/3}}{4\pi^2 m^*}$, where K_2 is therefore a constant. The metallic state for an electronic concentration (n_c) when $(E_c)_{N_e = n_c} > W_{el}$, so that:

$$K_2 \frac{h^2 (n^{2/3})_{N_e = n_c}}{4\pi^2 m^*} \geq \frac{e^2}{4\pi\varepsilon_0 K_1 a^*}$$

can also be written as:

$$a^* n_c^{2/3} \geq \frac{1}{K_1 K_2} \frac{\pi m^* e^2}{\varepsilon_0 h^2}$$

As $a^* = [\varepsilon_0 h^2]/[\pi m^* e^2]$, this condition can be written more simply as:

$$\left[a^* n_c^{1/3}\right]^2 \geq \frac{1}{K_1 K_2}$$

By making $C = [1 / K_1 K_2]^{1/2}$ we finally obtain the condition for the transition to the metallic state:

$$a^* n_c^{1/3} \geq C$$

Experimentally, the constant C is about eight orders greater than n_c. For example, an evaluation gave $C \approx 0.26$, and the transition is with $a^* = a_1(m/m^*)$ for the effective Bohr radius:

$$a^* n_c^{1/3} \geq 0.26$$

Physically speaking, this criterion means that all materials can become metallic if they are sufficiently compressed so that the electron density reaches the value n_c. The corresponding metal-insulator transition (*M-I* transition, which also occurs at $n = n_c$) is called the Mott transition and originates from localization of electrons through electrostatic interactions, not from any material disorder. We shall see in section 9.3 how disorder alone can result in the so-called Anderson transition.

9.2.4. *Additional material: examples of the existence and inexistence of Mott–Hubbard transitions*

9.2.4.1. *Charge transfer complexes*

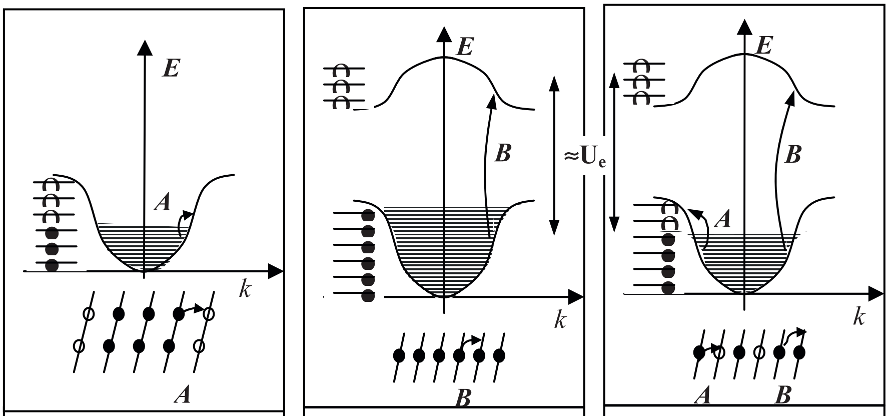
Charge transfer complexes (CTCs) are materials in which the effective correlation energy is high. In this case, the effective energy (U_{eff}) is defined as the difference between the electronic repulsion energy for a site occupied by two electrons (U_o) and the electronic repulsion energy between two electrons on adjacent sites (U_1), i.e. $U_{\text{eff}} = U_0 - U_1$.

For a system with N sites:

– If we can assume that U_{eff} is negligible, each site can be occupied by two electrons (spin up, ↑, and spin down, ↓). In addition, as in Figure 9.4a, if the system is half filled by N electrons then the material is metallic.

– If the system is one in which U_{eff} is high, we can place only one electron per site. Again, if the system carries N electrons (i.e. $\rho = 1$, in which ρ designates the number of electrons per site) then all energy levels are occupied and the band is full as shown in Figure 9.4b. Only B inter-band transitions are allowed, demanding a high energy of activation, E_a; the system, in other words, is an insulator (Mott insulator) or semiconductor. For example, the complex $HMTTF–TCNQF_4$, in regular columns, has $\rho = 1$, $E_a = 0.21$ eV with a room temperature conductivity $\sigma_{RT} = 10^{-4}\ \Omega^{-1}\ \text{cm}^{-1}$.

– Once again, if the system is one in which U_{eff} is high and we can only place one electron per site but $\rho < 1$ because bonds at the interior of each column are not fully occupied, both A intra- and B inter-band transitions are possible with the former requiring little activation energy and the latter a high activation energy. This is shown in Figure 9.4c. As an example, $TTF^{+0.59}–TCNQ^{-0.59}$ displays a metallic character with $\rho = 0.59$ and $\sigma_{RT} = 10^3\ \Omega^{-1}\ \text{cm}^{-1}$.



(a) $U_{\text{eff}} \approx 0$: electrons in 2 states, ↑and ↓, allowed per site. If system has *N* sites and *N* electrons, have semi-full band and a metallic state.

(b) U_{eff} high, and only 1 electronic state allowed per site. If system has *N* sites for *N* electrons ($\rho = 1$), implies full band and material is insulator or semiconductor and only inter-band *B* transitions are allowed.

(c) U_{eff} high, and only 1 electronic state allowed per site. If $\rho < 1$, band partially full and both *A* intra- and *B* inter-band transitions are possible.

Figure 9.4. *Electron transport with respect to electronic structure. Upper parts of the figures represent band schemes and lower parts represent electron positions (• = occupied state, o = empty state)*

9.2.4.2. *π-conjugated polymers*

Polymers conjugated by π-orbitals are, in principle, not subject to Mott transitions as transfers from one site to another in the same chain have β integral values which are too high (typically of the order of $4\beta \approx 10$ eV for polyacetylene), well above electron–electron interaction energies (*U*, below 1 eV for polyacetylene). Figure 9.4a therefore sufficiently describes these materials, although they do display insulating characteristics, which in the case of polyacetylene results from a Peierls distortion due to electron–phonon interactions which open the band gap (Figure 9.5).

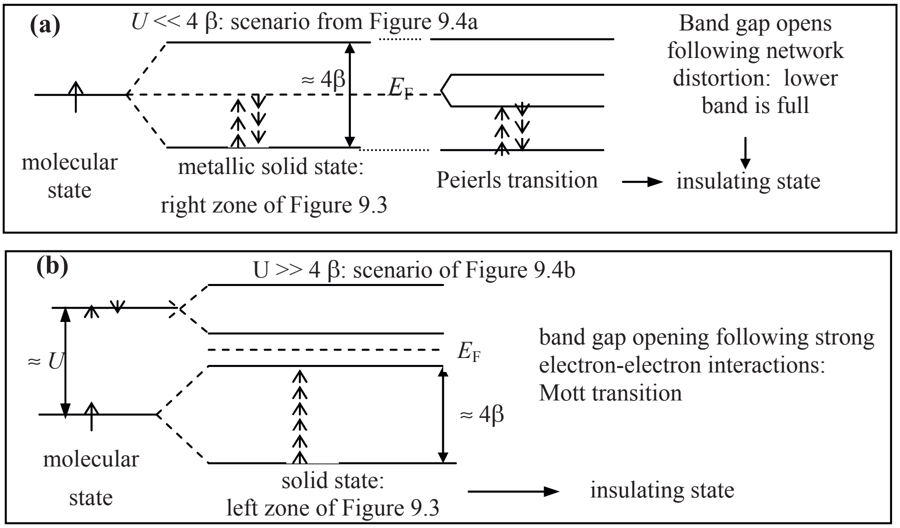


Figure 9.5. *(a) Characteristics of π-conjugated polymers, with the example of polyacetylene under a Peierls transition, verifying that U<<4β, in contrast to (b) Mott insulators for which U>>4β (and possible for CTCs)*

9.3. Effect of geometric disorder and the Anderson localization

9.3.1. *Introduction*

The effect of geometric disorder has for the most part been studied within theories on amorphous semiconductors developed by Mott and Davies (Mott and Davis, *Electronic Processes in Non-crystalline Materials,* Clarendon, 1979).

The theory is based on two fundamental ideas:

– The first idea was taken from the work of Ioffe and Regel who observed that there was no great discontinuity in the electronic properties of semi-metallic or vitreous materials when going from solid to liquid states. It was concluded that the electronic properties of a material cannot be only due to long-range order, as was proposed by Bloch for properties of crystals, but are also determined by atomic and short-range properties in which the average free path of an electron is inter-atomic. It is worth noting also that, even though a material may be amorphous, this does not exclude it from having bands. For example, glass, which is a non-crystalline material, is transparent in the visible region of light (≈1.5–3 eV), that is to say that while absorption of photons with energy below 3 eV does not occur, glass does actually have a band gap of greater than 3 eV.

– The second idea is based on the evidence given by Anderson for a material without long range order that nevertheless have localized states with permitted energy bands for electrons. This theoretical model comes from observations made on certain amorphous semiconductors in which charge carriers cannot move.

9.3.2. *Limits of band theory application and the Ioffe–Regel conditions*

Bloch functions, i.e. $\psi_k(\vec{r})$, can be used to describe electron wave functions in perfectly crystalline materials. The electronic states are delocalized and spread out over space, as denoted by $|\psi_k(\vec{r})|^2$. Because of perfect delocalization, the mean free path of an electron can be considered equal to infinity.

It is only when studying a real crystal though that the mean free path of an electron takes on significance because of effects due to quasi-imperfections caused by vibrations, called phonons (see Chapter 10), and imperfections caused for example by doping agents and impurities which perturb the regularity of potential throughout the network. It is only when these electron scattering effects, which limit the free path of electrons, are considered that the statistical average term ℓ of the free path length of an electron between two successive collisions can be introduced. In addition, there are two terms to note: “lattice scattering” which indicates collisions due to the material network; and for a similar effect caused by ionized impurities, the term “impurity scattering” is used.

On disordering a lattice by introducing vibrations and/or impurities, ℓ appears and takes on a decreasing value as disorder increases. If there is a low amount of impurities, then local levels appear, most notably in the forbidden band (FB), but if the number of impurities increases, the localized levels grow to form impurity bands which can reach a size DE_e, close to that of the valence band (VB), the conduction band (CB) and the FB introduced in Bloch’s model. Bloch's model however loses all semblance of reality when values of DE_e reach the same values of the bands.

Put another way, we can go from the crystalline state to the amorphous state with ℓ decreasing until Bloch’s theory is no longer acceptable. The limit for ℓ was fixed to $k\ell \sim 1$ (for a perfect crystal, $k\ell >> 1$) by Ioffe and Regel after following the reasoning of the uncertainty principle, i.e.:

$$\Delta E . \Delta t \geq \hbar \text{ and } \Delta x . \Delta k \geq 1 \qquad [9.1]$$

To arrive at the result shown above, we can consider that the trajectory of an electron after a collision is random, and at the very best can only be defined between two collisions, i.e.:

$$(\Delta t)_{\max} = \tau \qquad [9.2]$$

in which τ is the relaxation time – the average time between two collisions – and:

$$(\Delta x)_{\max} = \ell \qquad [9.3]$$

From equation [9.1] we can thus directly derive the best precision in ΔE, $(\Delta E)_{\min}$, and in Δk, $(\Delta \underline{k})_{\min}$, (1) when the equivalence of [9.1] is verified, as in $\Delta E\ \Delta t = \hbar$ and $\Delta x.\ \Delta k = 1$;, and (2) when Δt and Δx are at their highest value in equations just above and equal to $(\Delta t)_{\max} = \tau$ and at $(\Delta x)_{\max} = \ell$.

We then arrive at:

$$(\Delta E)_{\min}.\tau \approx \hbar \qquad [9.4]$$

and:

$$(\Delta k)_{\min}.\ \ell \approx 1 \qquad [9.5]$$

The question that we are therefore brought to ask is, with increasing disorder, what are the lowest values that τ (and thus the mobility $\mu = q\tau/m$) and ℓ can go to while $(\Delta E)_{\min}$ and $(\Delta k)_{\min}$ retain acceptable values, i.e. values which are compatible with classic band theory for real crystals.

The response can be given using simple calculations which show that:

– When $\mu \rightarrow 1$ cm^2 V^{-1} s^{-1} (and $\tau \approx 6 \times 10^{-16}$ s), from equation [9.4] $(\Delta E)_{\min} \approx 1$ eV. Thus, $(\Delta E)_{\min} \approx E_G$ (band gap size) or $(\Delta E)_{\min}$ is the same order of size as the permitted bands. When $\mu \leq 1$ cm^2 V^{-1} s^{-1}, the uncertainty in the energy of the carriers tends to the same order of size as the permitted and forbidden bands, to such an extent that the band scheme loses its relevance to real systems. It will be shown though that the Anderson model band scheme has to take into account localized bands with a gap which eventually becomes the mobility gap, E_μ.

– When $\ell \rightarrow$ a few ångström, that is to say $\ell \approx a$, in which a is the lattice constant and is typically of the order of 3×10^{-8} cm, equation [9.5] results in $(\Delta k)_{\min} \approx 1/\ell \approx 1/a \approx 3 \times 10^{7}$ cm$^{-1} \approx k$. In effect, with $\lambda = h/mv$ in which $v = v_{\text{thermal}} \approx 100$ km s^{-1} and $\lambda \approx 7 \times 10^{-7}$ cm, we have $k = 2\pi/\lambda \approx 10^{7}$ cm^{-1}. We can directly infer that in the band scheme, conduction electrons will be such that $k \approx 1/a$. At these values, where $\ell \approx a$, we thus have $(\Delta k)_{\min} \approx k$, and k can no longer be considered a good physical parameter to which we can apply quantification. In addition, when $\Delta k \sim k$, Fermi's sphere is badly defined (to the point of being totally deformed) and the concept of carrier speed looses significance as $\hbar k = m^*v$, just like the mean free path which is expressed as a function of v, i.e. $\ell = v\,\tau$.

Finally, as soon as $\ell \approx a$, and more strictly speaking as soon as $\ell \leq a$ that occurs when the interaction between an electron and the material network becomes increasingly strong, an electron no longer goes any further than the limits of the atom to which is tied. The electronic wave function localizes over a small region in space and is generally supposed to diminish exponentially with respect to R (as $\exp(-\alpha R)$).

The work of Mott and Anderson has introduced a new concept of localized states. The permitted density of states, $N(E)$, always results in an energy band beneath a single E_C for a conduction band and above a single E_V for a valence band, and, in other words, an activation energy is necessary for carriers to pass from one state to another with an emission or absorption of a phonon.

9.3.3. *Anderson localization*

9.3.3.1. *The model*

Systems in which disorder is due to a random variation in the energetic depth of regularly spaced sites (with interstitial distances always equal to a) are considered in Anderson’s model, and can relate, for example, to a random distribution of impurities. Different authors, including Mott, have tried to take into account lateral, spatial disorder and the results have been close to those of the Anderson model, to which we will limit our discussion.

In Chapters 1 and 3, we saw that if we take into account effects resulting from a network of atoms at nodes by constructing a regular distribution of identical potential wells, then a permitted energy band of height B appears, as shown in Figure 9.6.

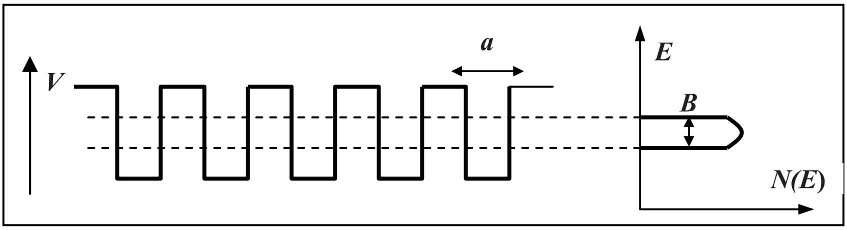


Figure 9.6. *Regular distribution of identical potential wells and permitted band*

Section 8.1 showed that in the strong bond approximation:

$-B = 2Z\beta$, where Z = the number of adjacent neighbors, and β is the resonance integral between two adjacent sites; and

$$-m^* = \frac{\hbar^2}{2\beta a^2} = \frac{\hbar^2}{a^2 B} Z \qquad [9.6]$$

so that $\mu = \frac{q\tau a^2}{\hbar^2}\frac{B}{Z}$, and semiconductors possessing a narrow B band exhibit low mobilities.

For Anderson's model we can replace the preceding distribution (Figure 9.6) by one of randomly deep potential wells which represent disorder, as shown in Figure 9.7.

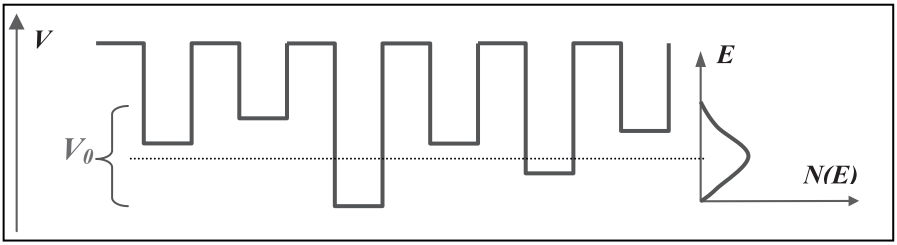


Figure 9.7. *Periodic distribution of wells with random depths*

9.3.3.2. *Variation in wave functions with respect to V_0/B (Anderson) and ℓ (Ioffe and Regel)*

We will show here how permitted energy bands change into localized states if V_0/B goes beyond its critical value. In order to do this, we need to look at the following scenarios:

– *Real crystal: V_0/B is very low and ℓ is high.* In this case, the wave function is given using Floquet's theorem (equation [7.19]) which can be written with the normalization constant:

$$\psi_k(r) = \sum_n e^{ikr_n}\psi_0(r - r_n) \qquad [9.7]$$

The mean free path can be estimated from the Born approximation, by realizing that the wave vector of an electron (k_n) changes to k_m once the electron has undergone a collision of probability P_{mn} and that $P_{nm} = \frac{1}{\tau} = \frac{v}{\ell}$, and in addition P_{nm} is given by Fermi's "golden rule": $P_{nm} = \frac{1}{4}\frac{2\pi}{\hbar}|\Omega|^2_{\text{ave}} N(E_m)$ (see Mott and Davis), where here $|\Omega|_{\text{ave}} = (V_0/2)$ and the relation is given for a unit volume. For conduction electrons, for which $\underline{E_m} \approx E_F$, spread throughout a volume $V = a^3$ with a distribution of wells with random depths but averaging around $V_0/2$, we here have:

$$\frac{1}{\ell} = \frac{P_{nm}}{v} = \frac{1}{4}\frac{2\pi}{\hbar}\left(\frac{V_0}{2}\right)^2 a^3 \frac{N(E_F)}{v} \qquad [9.8]$$

where $N(E_F)$ is the density of states at the Fermi level and v the velocity of an electron at the Fermi level. As ℓ is large, the system is almost a perfect crystal, and therefore it is possible to state that:

$$N(E) = \frac{4\pi(2m^*)^{3/2}E^{1/2}}{h^3} \quad \text{and} \quad v = \left(\frac{2E}{m^*}\right)^{1/2}$$

Using equation [9.6] for the effective mass, equation [9.8] gives $\frac{a}{\ell} = \frac{(V_0/\beta)^2}{32\,\pi}$, and with $B = 2Z\beta$:

$$\frac{a}{\ell} = \frac{(2ZV_0/B)^2}{32\,\pi} \qquad [9.9]$$

– *When $(V_0/B) \approx 1$ (lowest disorder) corresponding to $\ell \approx a$, i.e. threshold disorder*. When $\ell \approx a$, equation [9.9] written for a cubic system in which $Z = 6$ results in $(V_0/B) = 0.83 \approx 1$. At this point when $\ell \approx a$ (and $V_0 \approx B$), the disorder is such that $\Delta k \approx k$ (following Ioffe and Regel), and under such conditions, at each collision, k randomly varies by Δk, the closest neighbor to k. In going from one potential well to the next, the wave function in equation [9.7] randomly changes and, according to Mott, loses its phase memory. It should thus be written in the approximate form:

$$\psi_k(r) = \sum_n A_n \psi_0(r - r_n)$$

with $A_n = c_n \exp(i\varphi_n)$, where A_n is a function with a random phase and practically constant amplitude.

Moreover, this amplitude is more constant than the variation between depths of neighboring potential wells, i.e. V_0 is low in Figure 9.7. In a model using two wells with potential depths V_1 and V_2 (as in Miller and Abrahams, 1960) the resulting wave function can take on either a symmetric or antisymmetric form, respectively, $\psi_S = A_1\psi_1 + B\psi_2$ or $\psi_A = A_1\psi_1 - B\psi_2$. We can therefore show that when $|V_1 - V_2| \ll |\beta|$ (i.e. V_0 is low), so $A_1 \approx A_2$, the difference in energy $(E_1 - E_2)$ between the two possible states is such that $|E_1 - E_2| \approx 2|\beta|$. A representation of this function is shown in Figure 9.8a for a network of several potential wells.

– When $(V_0/B) > 1$ (V_0/B) (just above threshold value): initial delocalization and medium disorder. In a system which corresponds to a great increase in disorder, and for the model of just two wells corresponding to an increase in the depth between the wells as in $|V_1 - V_2| = V_0$, the difference in energy, $|E_1 - E_2|$, increases to a corresponding level and A_1 differs from A_2. The amplitudes of the functions are no longer constant and the wave function displays increasing disorder both in amplitude and in phase (Figure 9.8b).

– System in which $(V_0/B) \gg 1$ ((V_0/B) well above threshold value): considerable delocalization and disorder. In this system a highly localized state is formed, as shown in Figure 9.8c, and as V_0 increases the localization is accentuated. In addition, there is no longer propagation along a line of potential wells and states are thus localized. An exponential decrease in the wave function starts to appear and is increasingly noticeable with increasing values of V_0. The wave function takes the form:

$$\psi(r) = \left[\sum_n A_n \psi_0(r - r_n)\right] e^{-\alpha r}$$

from which it is possible to write:

$$\psi(r) = \left[\sum_n A_n \psi_0 (r - r_n) \right] e^{-\frac{r}{\xi}} \qquad [9.10]$$

where ξ is the localization length.

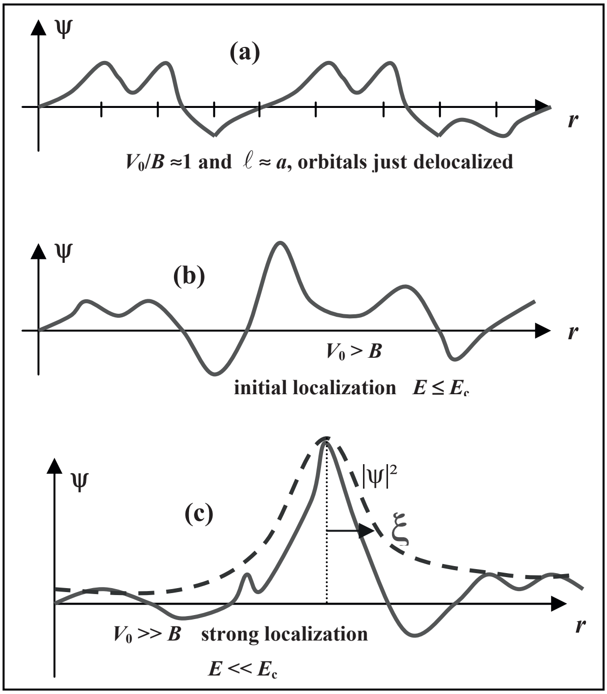


Figure 9.8. *Variations in wave functions with delocalization: (a) delocalization–localization only; (b) weak delocalization; and (c) strong delocalization*

To conclude, the factor V_0/B is a crucial term in deciding whether only localized states form ($V_0/B > 1$) or whether both localized and delocalized states can co-exist ($V_0/B \leq 1$).

9.3.3.3. *Band scheme and form of the state density function N(E)*

From a realistic scheme of the distribution of potential wells, we can see that states should be localized within one energy domain and delocalized in another. Accordingly, Figure 9.9 describes a system with non-negligible disorder:

– all states at the tail end of the function $N(E)$ which correspond to a high enough value of V_0 and from energies $E < E_c$ and $E > E'_c$ appear localized as before in the scheme of potential wells; and

– however, the middle of the band corresponds to shallow states with small V_0, such as V_0/B, and is a zone of delocalized states which have $E'_c < E < E_c$.

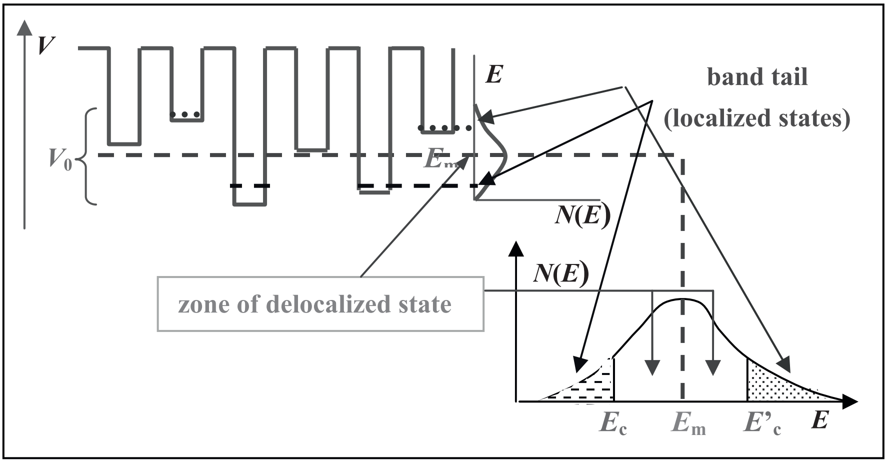


Figure 9.9. *Representation of localized and delocalized states in co-existence*

9.3.4. *Localized states and conductivity. The Anderson metal-insulator transition*

9.3.4.1. *Mott's definition*

Mott's definition is based on continuous conductivity relative to electrons with a given energy ($\sigma_E(0)$) and delocalized states are on average, at $T = 0\ K$, those for which $\sigma_E(0)$ is zero, i.e. $\langle \sigma_E(0) \rangle = 0$. To arrive at an average though, all possible configurations which have the energy E need to be considered, and while some electrons may have a non-zero energy, the average over all possible states with the corresponding energy E gives zero as a result. These states and the mobility they represent are in effect thermally activated.

However, at $T = 0\ K$, delocalized states average to give $\sigma_E(0) \neq 0$, that is to say metallic behavior occurs.

9.3.4.2. *State properties*

In Figure 9.9, two types of states – localized and delocalized – are separated by energies E_c and E'_c which together are called the mobility edge. In the two zones Einstein's relation holds true if E_F is outside the bands of non-degenerate states. We thus have $\mu = qD/kT$, but the diffusion coefficients (D) have different forms. Here, $D = Pa^2$ in which P represents the probability of movement to neighboring sites. This brings us to the origin of the expressions used in courses on electronics and materials:

– When $E > E'c$ and $E < Ec$, $D = (1/6)\ \nu_{ph}\ a^2 \exp(-w_1/kT)$ (with ν_{ph} being the phonon frequency and w_1 the energy of activation) and $<\sigma_E(0)>_{T=0\text{K}} = 0$. Here as $T \to 0$, we can verify that D and μ tend towards zero, as does the conductivity.

– When $Ec < E < E'c$, $D= (1/6)\ \nu_e\, a^2$ and $\sigma_E(0) \neq 0$ where ν_e is the frequency of electronic vibrations.

9.3.4.3. *Distinction between insulator (or semiconductor (SC)) and a metal in a slightly disorder material (poor localization and small ℓ)*

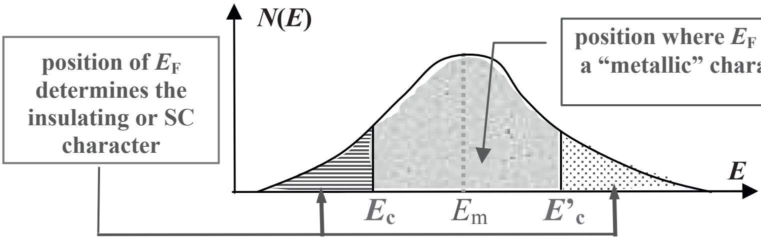


Figure 9.10. *Metallic character resulting from the position of E_F*

As for classic (crystalline) materials, the properties are tied to the position of E_F (as shown in Figure 9.10):

– when E_F is situated in the domain of delocalized states, ($E_c < E_F < E'_c$), we have a degeneration corresponding to a metallic character; and

– however, when E_F is situated in the zone of localized states, where typically $E < E_c$, charge carriers can only be thermally excited and conductivity can occur only by jumps or by excitation to E_c, and indeed at 0 K conductivity tends towards 0 which is typical of an insulator. Materials for which the Fermi level is situated in an energy zone in which states are localized are called Fermi glasses.

9.3.4.4. *Transition from metal to insulator or semiconductor*

For a given material which has a Fermi level fixed by its charge density, displacement of E_C, for example by increasing the disorder as shown in Figure 9.11, moves the Fermi level from an initial state in a domain of delocalized states (metallic) to a zone of localized states. The result is a metal to insulator or semiconductor transition.

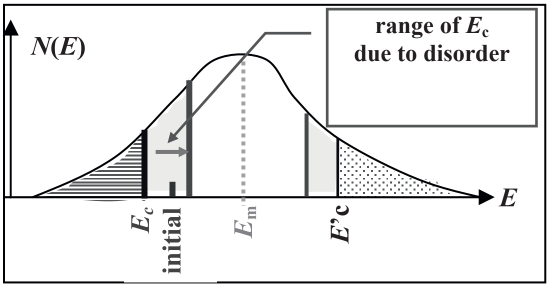


Figure 9.11. *Using disorder to displace E_C and effect the metal-insulator transition*

9.3.4.5. *Anderson transition from order to disorder and the change in conductivity*

Even though we do not detail transport properties in this book, we should nevertheless introduce an expression for metallic conductivity written in the relatively simple form of $\sigma = qn\mu = nq^2\tau / m^*$, in which n is electron concentration and τ is the relaxation time relative to the Fermi level.

With $\ell = v\tau$ we have $\sigma = nq^2 \ell / m^* v$, and on introducing the crystalline momentum, $\hbar k = m^* v$, we reach $\sigma = nq^2 \ell / \hbar k_F$ in which k_F is the wave vector at the Fermi surface.

We can also note that the number n of electrons within a unit volume V (V=1) can be obtained by use of the reciprocal space, that is to say the number of cells within the Fermi volume is $([4/3]\pi k_f^3) / 8\pi^3$, each with volume $8\pi^3/V = 8\pi^3$ for V =1.

On taking into account electron spins (i.e. doubly occupied cells), we have $n = 2([4/3]\pi k_f^3) / 8\pi^3$ and metallic conductivity can therefore be written as:

$$\sigma_B = \frac{4\pi k_F^2 q^2 \ell}{12\pi^3 \hbar} \qquad [9.11]$$

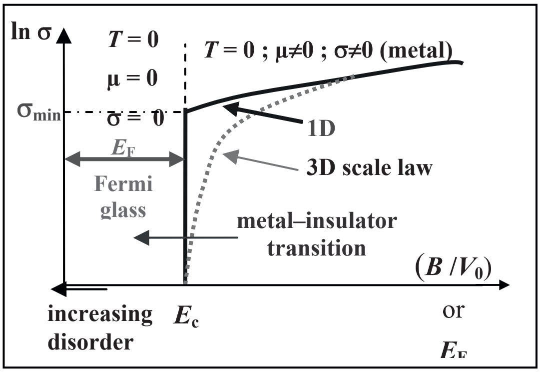


Figure 9.12. *Anderson's transition from metal to insulator at absolute zero following $1/(V_0/B) = B/V_0$. The same behavior occurs when E_F is displaced with respect to E_C*

When considering a metallic state, the Fermi level can be considered more or less in the middle of the band (so that in 1D, $k_F \propto \pi/a$) as shown in Figure 9.10. With increasing disorder, E_c and E'_c tend towards each other at the band center E_m ($\approx E_F$) in which case all states are localized. This change is called Anderson's transition, as shown in Figure 9.12.

At the same time, we have $V_0/B \geq 1$ with ℓ tending towards a. For its part, with $\ell = a$, conductivity σ thus tends towards $\sigma_{min} = \sigma_{IR} = (\sigma_B)_{\ell=a} = q^2/3a\hbar$. In mono-dimensional media this abrupt transition is a point of controversy as it is known to occur progressively in 3D. When a is of the order of 3 Å, $\sigma_{IR} = 700$ S cm^{-1}, it is often a value of saturation for conductivity in rising temperatures.

9.4. Conclusion

This chapter has shown how supplementary effects, defaults in the periodicity, caused by such as dangling bonds, chain-ends and holes within the structure introduced into the band gap localized levels can open to form a band when their fluctuation is brought into account. This band can split as a function of electron filling, in a manner analogous to the perturbations caused by electron repulsions, which were not taken into account in the band theory. These electronic repulsions were looked at in detail and we have shown how they can help to explain metallic or insulating behaviors that the classic band theory cannot deal with.

By introducing modifications of crystal regularity by considering network thermal vibrations (phonons) and defects, both chemical (impurities) and physical (dislocations), the notion of a real crystal was studied. This resulted in determining the mean free pathway of electrons, which can no longer be considered as completely delocalized within the network – as in the case of a perfect crystal. It was shown that an increase in disorder reduced the mean free path length up to the point of localizing electrons in neighborhoods of deep defaults, resulting in energy levels localized at extremities, or "tails", of permitted bands.

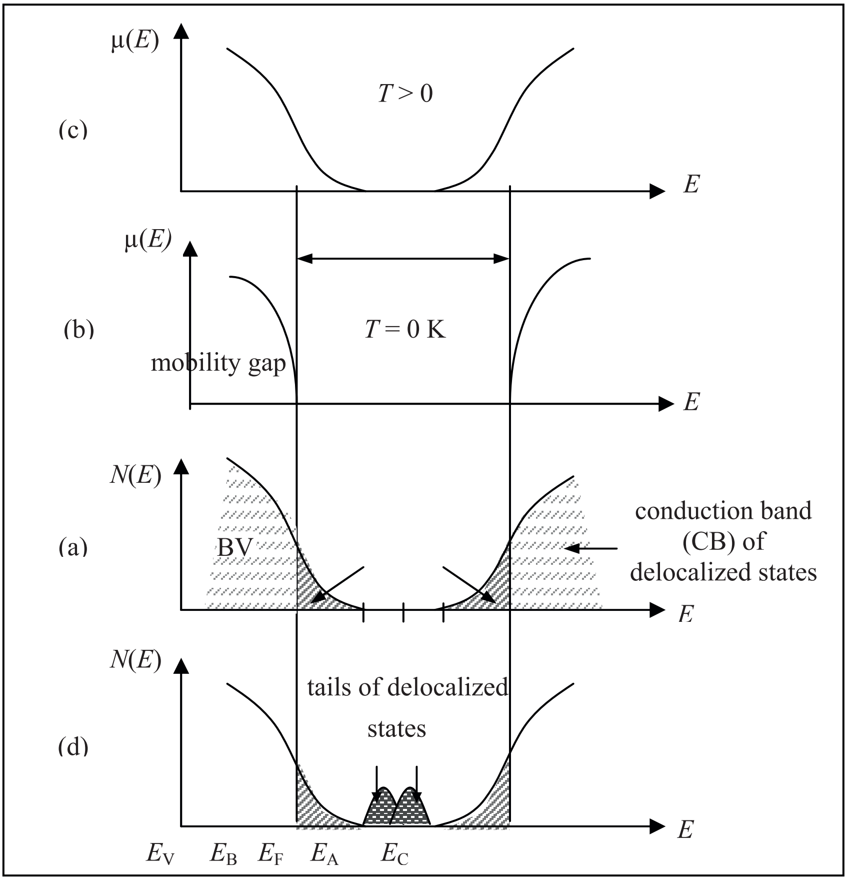


Figure 9.13. *Band models for amorphous semiconductors: (a) following CFO; (b) following $\mu(E)$ at $T = 0$ K; (c) following $\mu(E)$ at $T > 0$ K; (d) following Mott and Davis*

Finally, it can be noted that all models postulate for amorphous media, such as crystalline semiconductors, that there are conduction and valence bands which are or are not separated by a band gap, depending on the all-important band tails. And that:

– Bands result in part from short-range order (from approximations of strong bonds giving rise to bonding and anti-bonding states, i.e. valence and conduction bands separated by a band gap) and from disorder created by phonons or impurities shown by tails of delocalized states. Tail states are neutral when occupied in the case of the valence band and when empty in the conduction band. The Fermi level is thus placed in the middle of the band gap, as shown in Figure 9.13a and following the model proposed by Cohen, Fritzsche and Ovshinsky (Frizsche, 1970).

– The form of the bands depends on the type of the implicated orbitals. For *p* or *d* orbitals, which are less stretched overall into space than *s*-orbitals, the form of $N(E)$ is different and the bands are smaller.

– In a perfect crystal, the band gap is an forbidden energy in which $N(E) = 0$, while in an amorphous material it is a mobility gap and $N(E)$ is not necessarily zero but the mobility $\mu(E)$ however does becomes zero at $T = 0\ K$ (localized states), as shown in Figures 9.13b and c).

By taking into account the disorder caused by not only phonons and impurities but also by structural defects such as dangling bonds and chain ends, additional offsetting defaults localized in the middle of the band can generate two bands at compensating levels (Hubbard’s bands) following the model of Mott and Davis as shown in Figure 9.13d.

9.5. Problems

9.5.1. *Additional information and problem 1 on the Mott transition: insulator–metal transition in phosphorus doped silicon*

9.5.1.1. *Additional information: evidence for the Mott transition in phosphorus doped silicon*

The Mott transition can be illustrated by way of the conductivity of silicon at low temperatures as a function of the concentration of phosphorus atoms (see Figure 9.14).

Substituting silicon atoms, four or five electrons in the external phosphorus layer are engaged in covalent bonds with neighboring silicon atoms; the fifth electron dopes the silicon. When $T = 300\ K$, we have $kT \approx 0.024$ V so that the bonding energy of this fifth electron is of the order of 0.045 eV and the P atoms are easily ionized to P^+ at ambient temperatures. The phosphorus plays the role of a donor for *n*-type doping of silicon.

At very low temperatures though, kT is very small as the phosphorus electrons are tied to the P^+. The fifth electron therefore tends to follow a Bohr radius calculated in section 9.2, as in $a^* \approx \varepsilon_r\, a_1(m/m^*)$, which in silicon is given by

$a^* \simeq 25$ Å. The material is therefore an insulator in the domain of very low temperatures, at least when concentrations of phosphorus are below a critical point (found experimentally to be around 3.8×10^{18} P atom cm^{-3}). An abrupt insulator–metal transition is found around this concentration.

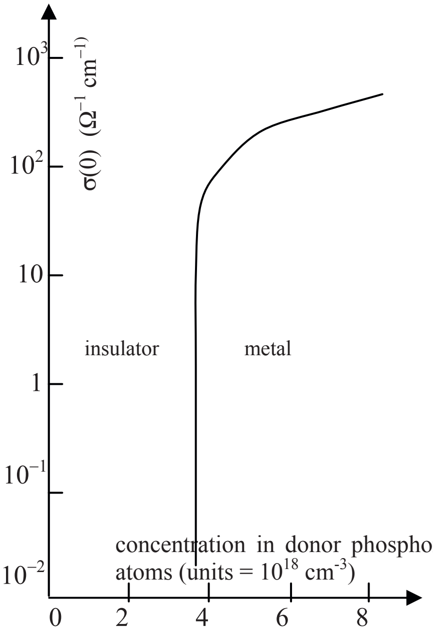


Figure 9.14. *Conductivity of silicon doped with phosphorus near absolute zero*

It is possible to consider that at high concentrations of phosphorus atoms, their orbitals overlap so that electrons can pass by tunneling from one P^+ ion to another. The resulting electronic cloud can play a role as electrostatic screen between the P+ and the fifth electron so that the latter end up being delocalized in the cloud. Once this process of liberation (with respect to the atom of origin) has occurred, it gains strength through a growing screening effect.

9.5.1.2. *Problem: the Mott transition as an equation and its resolution*

The mechanism for the Mott transition can be characterized physically in the following manner: if a nucleus with a positive charge (denoted here Ze) is placed in a free electron gas, the electrons will be attracted to the positive charge and generate in the neighborhood of the nucleus a surplus negative charge density. This negative space charge will then screen the positive charge and the potential generated by the nucleus ($V_0 = \frac{Ze}{4\pi\varepsilon_0 r}$) in the vacuum will be reduced so that in the material,

$V = \frac{Ze}{4\pi\varepsilon_0\varepsilon_r r}$. Parameter ε_r is introduced to account for the nature of the material. If it is a dielectric, the electronic charges are tied by covalent bonds to the cores of the ions and are, as a consequence, not very mobile. The "screening" is therefore poor and ε_r is small (typically of the order of two or three). If the medium is metallic, the electrons are very mobile and the screening effect can be complete some distance from the nuclei. In such cases, $V \to 0$ and therefore $\varepsilon_r \to \infty$ (total "screening").

Nevertheless, this screen is effective right from the very shortest distance, of the order of an ångström. At greater distances the electrons no longer see the potential of the nuclei and it is for this reason that electrons behave independently from the nuclei at distance greater than the screen. It is this distance that we will try to evaluate.

The most mathematical treatment of the screen effect can be carried out within the Thomas–Fermi approximation. It is assumed that the local perturbation created by the electronic cloud has a potential δV (generated by the positive charge) and reduced locally by the energies of the electrons by $dE = |e|\delta V$. The result is that the electronic density function is modified by δn.

1) Express δn using the density of states function in E_F (function $Z(E_F)$).

2) Write the Poisson equation that ties δn and δV at a point far from the positive cores.

3) Assuming that the medium is isotropic, the perturbation potential can be thought of as being constant for a given r, i.e. it has a spherical symmetry. Write the Poisson equation for these conditions.

4) Give the physical solution to this differential equation by introducing the parameter $\lambda_{TF} = \left[\frac{e^2}{\varepsilon_0} Z(E_F)\right]^{-1/2}$ which is the Thomas–Fermi screening length.

5) Determine the solution using the limiting condition that δV tends towards the Coulombic potential as $\lambda_{TF} \to \infty$. Make a comparison using a graphical representation.

6) Give the equation for the spatial variation in the dielectric permittivity.

7) Express λ_{TF} as a function of the electronic concentration (N_e) and the parameter $a^* = a_1(m/m^*)$ (where $a_1 = [\varepsilon_0 h^2]/[\pi m e^2]$).

8) What relation should one have between a^* and λ_{TF} for the medium to be considered metallic? From this deduce the relation that gives the critical electronic concentration (n_c) above which the medium is metallic.

Answers

1) If a local perturbation δV (by an electronic cloud) of the potential reduces the potential energy of the electrons by $dE = |e|\delta V$, the electronic density is modified by $\delta n = Z(E_F)\, dE = Z(E_F)\, |e|\delta V$, where $Z(E_F)$ is the density of states function (or state density function – the term is interchanged in this book) at the Fermi level (E_F), the level at which the conduction electrons are transported.

2) Far from the cores of the positive charges, where the volume density of the charge is $\rho = -e\,\delta n$, the Poisson equation that relates δn and δV is $\Delta(\delta V) + \frac{-e\delta n}{\varepsilon_0} = 0$ (with the medium of the material being represented by charge that bath in a vacuum, the permittivity being used is thus ε_0). With $\delta n = Z(E_F)\, |e|\delta V$ we thus find:

$$\Delta(\delta V) - \frac{e^2}{\varepsilon_0} Z(E_F)\delta V = 0$$

3) Supposing that the medium is isotropic, the potential due to the perturbation (δV) can be assumed to be constant for a given r, that is within a spherical symmetry. In this case, the Laplacian is in the form:

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial(\delta V)}{\partial r}\right) = \frac{\partial^2(\delta V)}{\partial r^2} + \frac{2}{r}\frac{\partial(\delta V)}{\partial r} = \frac{1}{r}\frac{\partial^2\left[r(\delta V)\right]}{\partial r^2}$$

so that the Poisson equation takes the following form:

$$\frac{1}{r}\frac{\partial^2\left[r(\delta V)\right]}{\partial r^2} - \frac{e^2}{\varepsilon_0} Z(E_F)\delta V = 0$$

which can then be written in the form of a classic differential equation:

$$\frac{\partial^2\left[r(\delta V)\right]}{\partial r^2} - \frac{e^2}{\varepsilon_0} Z(E_F)\left[r(\delta V)\right] = 0$$

4) The solution for this differential equation is:

$$[r(\delta V)] = Ae^{-\left[\frac{e^2}{\varepsilon_0}Z(E_F)\right]^{1/2} r} + BAe^{+\left[\frac{e^2}{\varepsilon_0}Z(E_F)\right]^{1/2} r}$$

so that:

$$(\delta V) = A\frac{e^{-\left[\frac{e^2}{\varepsilon_0}Z(E_F)\right]^{1/2} r}}{r} + B\frac{e^{+\left[\frac{e^2}{\varepsilon_0}Z(E_F)\right]^{1/2} r}}{r}.$$

As:

$$\lim_{x \to +\infty}\left(\frac{e^x}{x^m}\right) = +\infty,$$

with $m > 0$, the second term diverges and we have to take $B = 0$, so that:

$$\delta V = A\frac{e^{-\left[\frac{e^2}{\varepsilon_0}Z(E_F)\right]^{1/2} r}}{r}.$$

By making the hypothesis, $\lambda_{TF} = \left[\frac{e^2}{\varepsilon_0}Z(E_F)\right]^{-1/2}$ for the Thomas–Fermi screening length we thus find $\delta V = A\frac{e^{-\frac{r}{\lambda_{TF}}}}{r}$.

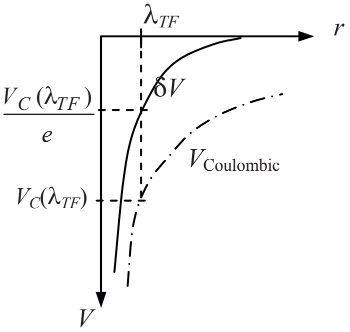
5) In the exponential, if r → 0, or what comes to the same thing, if $\lambda_{TF} \to +\infty$, for an infinitely long screen, the potential must tend towards that produced by one single atomic core, that is its normal Coulombic potential, such that:

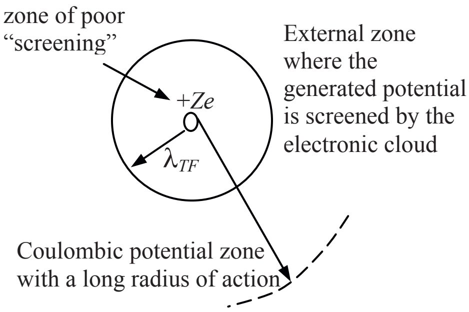
$$A\frac{e^{-\frac{r}{\infty}}}{r} \to \frac{A}{r} = \frac{Ze}{4\pi\varepsilon_0 r}, \text{ from which } A = \frac{Ze}{4\pi\varepsilon_0}.$$

Finally, we have:

$$\delta V = Ze\frac{e^{-\frac{r}{\lambda_{TF}}}}{4\pi\varepsilon_0 r}.$$

The electrostatic potential therefore decreases much more rapidly with distance than if it were simply the Coulombic potential.





6) In a representation of the medium with the help of the dielectric permittivity, $\varepsilon_0\varepsilon_r$, where ε_r is the relative dielectric permittivity, and takes into account the polarization of the medium, the potential generated by the charge *Ze* is given by $\frac{Ze}{4\pi\varepsilon_0\varepsilon_r r}$. Identification with the preceding form of the equation, $\delta V = Ze\frac{e^{-\frac{r}{\lambda_{TF}}}}{4\pi\varepsilon_0 r}$, gives $\varepsilon_r = e^{\frac{r}{\lambda_{TF}}}$. It is possible to see that when $r >> \lambda_{TF}$, $\varepsilon_r \rightarrow \infty$.

This result is normal as in this region the electrons are not subject to any influence from the charge *Ze*, which would tend to localize them. When $r >> \lambda_{TF}$ the electrons are free as in a metal, for which the permittivity also tends towards infinity.

7) In addition, $Z(E_F) = \frac{4\pi(2m^*)^{3/2}}{h^3}\sqrt{E_F} = A_c\sqrt{E_F}$, where $A_c = \frac{4\pi(2m^*)^{3/2}}{h^3} = 4\pi\left(\frac{2m^*}{h^2}\right)^{3/2}$, and where E_F can be expressed as a function of the electronic density by writing, for example, that at absolute zero (see problem 4 in Chapter 2):

$$N_e = \int_0^{E_F} Z(E)dE = A_c \int_0^{E_F} E^{1/2}dE = \frac{2}{3}A_c E_F^{3/2} \text{ so that } E_F = \left(\frac{3}{2}\frac{N_e}{A_c}\right)^{2/3},$$

from which

$$Z(E_F) = A_c^{2/3}\left(\frac{3}{2}N_e\right)^{1/3}.$$

The result is that:

$$\lambda_{TF} = \frac{(\varepsilon_0)^{1/2}}{e}\left[\frac{1}{A_c}\right]^{1/3}\left[\frac{2}{3}\frac{1}{N_e}\right]^{1/6} = \frac{1}{2}\sqrt{\frac{\varepsilon_0 h^2}{\pi^{2/3}e^2 m^*}}\left(3N_e\right)^{-1/6},$$

so that with $a_1 = [\varepsilon_0 h^2]/[\pi m e^2]$ and $a^* = a_1(m/m^*)$:

$$\lambda_{TF} = \frac{a^{*1/2}}{2}\left[\frac{\pi}{3N_e}\right]^{1/6} \simeq 0.5\left[\frac{a^{*3}}{N_e}\right]^{1/6}.$$

8) The condition for electronic behavior is that a^* (which gives the position of the electron in its orbit) is greater than the screening length. This is because above this distance, the electron is no longer held by its starting atom. The result is the condition $a^* \geq \lambda_{TF}$.

This condition is fulfilled when N_e attains a critical concentration $(n_c)a^* \geq \frac{1}{2}\frac{(a^*)^{1/2}}{n_c^{1/3}}$, so that:

$$n_c^{1/3}a^* \geq 0.25$$

in a condition practically identical to that found in section 9.2.3.

9.5.2. *Problem 2: transport via states outside of permitted bands in low mobility media*

1) What is the condition that the width of the *B* band of a material must verify for the effective mass approximation to be valid when used in an expression for mobility?

2) Considering a system where the permitted bands are wide (such as in π-conjugated polymers), of the order of 1 eV, and where the mobility is of the order of 10^{-3} cm² V^{-1} s^{-1}. Can electronic transport occur in the permitted bands?

3) We now consider solids as molecular solids where intermolecular bonds are through weak van der Waals bonds. Can it be said that this system has a narrow band? In this example where the mobility can be taken as being in a general form (for a 1D system), $\mu = \frac{q\ell}{kT}\langle v_x \rangle$, and where it may reach 1 cm² V^{-1} s^{-1}, does the transport occur through these narrow permitted bands? Using the example of anthracene, the intermolecular distance is given by $a \approx 6 \times 10^{-8}$ cm. For cases where we know the resonance integral ($\beta \approx 0.01$ eV for anthracene), give an alternative rationale.

Answers

1) To define the effective mass, as in section 8.1, the width (*B*) of the permitted bands must be such that $B >> kT$, or otherwise the band width will approach kT (0.0026 eV at ambient temperature), and it will not be just the lowest levels of the band that are occupied but all possible levels that can be taken up through thermal agitation. The form of the mobility for a strong bond is no longer acceptable. Expression [8.11] from Chapter 8, $\mu = \frac{q\tau a^2}{\hbar^2}\frac{B}{Z}$, obtained with the help of the classic expression for mobility $\mu = \frac{q\tau}{m^*}$, in which is introduced the effective mass

$m^* = \frac{\hbar^2}{Ba^2}Z$ (see equation [8.11’] where Z is the coordination number). In effect, to obtain this expression for m^*, we use an approximation that excludes the states at the base of the band ($k \approx 0$, and situated at the center of the zone for which E is a minimum, so that $\left[\frac{\partial E}{\partial k}\right]_{k \approx 0} = 0$ also allows the limited development over cosines which is involved in the expression for a strong bond).

2) From the preceding section the approximation of the effective mass is valid when considering wide bands, for example HOMO and LUMO bands in π-conjugated polymers. Inside these bands, the carrier mobility can be evaluated with the help of the equation $\mu = \frac{q\tau a^2}{\hbar^2}\frac{B}{Z}$.

In addition, and as has been detailed in this chapter (see section 9.3.2), for bands to retain a physical significance $B > \Delta E$, where $\Delta E \approx \frac{\hbar}{\tau}$ (equation [9.4]). Under these conditions for there to be conduction in these large bands of delocalized states, we should also verify that:

$$\mu = \frac{q\tau a^2}{\hbar^2}\frac{B}{Z} > \frac{q\tau a^2}{\hbar^2}\frac{\Delta E}{Z} \approx \frac{qa^2}{\hbar}\frac{1}{Z}.$$

This equation, $\mu > \frac{qa^2}{\hbar}\frac{1}{Z}$, is the condition for conduction in the B bands of delocalized states. With a ≈ several Å (length of strong bonds in π-conjugated polymers) and $Z \approx 2$, we find the condition: $\mu > 10^{-1}$ cm² V-1 s^{-1}. As the mobilities in π-conjugated polymers are of the order of 10^{-3} cm² V^{-1} s^{-1}, we can conclude that it is not reasonable to expect the transport observed in these materials to be in the bands of delocalized states. As a consequence, we can think that in such polymers, the mobility associated with delocalized states is greatly reduced by charge transport through more localized states.

3) Molecular solids give rise to narrow permitted bands, as the intermolecular bonds are through weak bonds, for example van der Waals bonds. The poor overlap of intermolecular orbitals results in low values for the overlap integrals (β). The width (B) of the permitted bands being of the order of several β ($B = 2Z\beta$, see equation [8.10]), means that the system is based on narrow permitted bands. In this case, and taking the argument developed in answer 1 into account, the approximation for the effective mass used to evaluate the mobility is no longer acceptable.

We can therefore use the following expression for mobility: $\mu = \frac{q\ell}{kT}\left\langle v_x \right\rangle$. As the speed can be given (in a dualistic theory) by $v = v_g = \frac{\partial\omega}{\partial k}$, so that with $E = \hbar\omega$ and in 1D we have $v = \frac{1}{\hbar}\frac{\Delta E}{\Delta k_x}$. With $\Delta E \approx B$ while $\Delta k \approx 1/a$ (order of size of the Brillouin zone in the reciprocal space and corresponding to the height of the permitted band (B) (see the representation of $E = f(k)$ in Chapter 4) we thus have:

$$v_x \approx \frac{Ba}{\hbar}, \text{ so that } \mu = \frac{q\ell}{kT}\frac{Ba}{\hbar}.$$

At this level we can reason in one of two ways (which lead, of course, to similar conclusions). Either:

– we consider that to have conduction in the localized states, we should have $\ell > a$ (Ioffe and Regel's second condition – see section 9.3.2), which means that:

$$\mu > \frac{qa^2}{kT}\frac{B}{\hbar} \text{ must be true.}$$

In these small molecules, with the band width $B \approx kT$, we find the same condition as in polymers:

$$\mu > \frac{qa^2}{\hbar}.$$

Taking a $\approx 5 \times 10^{-8}$ cm, we should have $\mu \geq 1$ to 10 cm² V^{-1} s^{-1}. It should be noted that here *a* represents the intermolecular distances that are slightly longer than the covalent bonds , so that we can reasonably state that a^2 for small molecules is at least an order greater than for polymers. With $\mu \approx 1$ cm² V^{-1} s^{-1} and $a \approx 6\times10^{-8}$ cm for anthracene, we do not find that the inequality is verified, which shows $\mu \geq 5$ cm² V^{-1} s^{-1}; or

– we consider that we know the mobility ($\mu \approx 1$ cm² V^{-1} s^{-1} for anthracene) and given that $B \approx \beta$ (where the resonance integral $\beta \approx 0.01$ eV for anthracene), and with $a \approx 6\times10^{-8}$ cm, we can therefore estimate ℓ with the help of $\mu = \frac{q\ell}{kT}\frac{Ba}{\hbar}$, which gives us $\ell \approx 3\times10^{-8}$ cm. The mean free pathway appears to be less than the intermolecular distance *a*, which again is not compatible with conduction in delocalized states (Ioffe and Regel).

Chapter 10

The Principal Quasi-Particles in Material Physics

10.1. Introduction

In a physical system there are a variety of particles that exist under constant perturbation, even excitation, due to external forces that are mostly interacting with the lattice. A slightly perturbed or excited state can dissipate into elementary perturbations (excitations) that propagate through the lattice just like a particle with a degree of energy, movement, and even a spin moment. A change in the perturbed (excited) state of a material can be described using variations in these parameters of energy, movement and so on. And these variations can be due to collisions between quasi-particles that can also be described using the same parameters. These quasi-particles are the result of interactions (perturbations and excitations) between real particles or the same particles with the lattice.

Depending on the nature of the interactions, the quasi-particles can be, for example:

– phonons: these describe the state of a lattice in which the atoms are excited by thermal vibrations. The main applications are in the domains of materials and thermal properties;

– polarons: these appear when accounting for coupling between electronic charges and the resulting lattice deformations. They are well evident in descriptions of transport phenomena in materials with low mobilities;

– excitons: these are coupled electron-hole pairs with energy levels in the neighborhood (but outside) of the permitted bands. They are essential to descriptions of the optical properties of materials;

– plasmons: these detail the collective oscillations occurring in an electron gas.

There are other quasi-particles that can be present. These include polaritons, that describe the transverse optical coupled modes of photons and phonons, magnons, that describe systems based on coupled spin moments, and solitons that correspond to a default in the conformation of a chain of atoms that may or may not be coupled to a charge. However, this text will detail only the main types of quasi-particles.

10.2. Lattice vibrations: phonons

10.2.1. *Introduction*

With the bonding forces between atoms being finite, an external perturbation such as thermal energy can result in atoms being distanced from the equilibrium position.

In this example the forces and atomic displacements are directed along r. With $F(r)$ denoting the bonding force on an atom at r with the corresponding potential energy $W(r)$, then:

$$F(r) = -\text{grad}\,W(r) = -\frac{dW}{dr}$$

If r_0 represents the equilibrium position of an atom (located with respect to the origin of the displacements), and if Δr denotes the displacement that is such that $r = r_0 + \Delta r$, we can write that:

$$W(r) \simeq W(r_0) + \frac{1}{2}\left(\frac{d^2W}{dr^2}\right)_{r_0} (\Delta r)^2$$

(as the first derivative is zero with respect to r_0 where $W(r)$ is a minimum and an equilibrium position). The result is:

$$F(r) = -\frac{dW(r)}{dr} = -\left(\frac{d^2W}{dr^2}\right)_{r_0} \Delta r = -k\Delta r, \text{ when } k = \left(\frac{d^2W}{dr^2}\right)_{r_0}$$

and we have Hooke's law. With $dr = d\left(\Delta r\right)$, the fundamental dynamic equation is written as:

$$m\frac{d^2\left(\Delta r\right)}{dt^2} = -k\Delta r$$

for which the physical solution is of the form:

$$\Delta r = A\cos\sqrt{\frac{k}{m}}t = A\cos 2\pi\nu t$$

where $\nu = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$. This represents the sinusoidal oscillatory movement with frequency ν.

In a crystalline system it is no longer a single atom that vibrates but rather a set of atoms and a coupling between the vibrations. This is because the vibrations of atoms will affect those of neighboring atoms.

If we have a system with N atoms, and each atom has three degrees of freedom, then the vibrational system will have $3N$ waves with different frequencies (normal vibrations). The movement of each atom is determined by the superposition of the $3N$ waves (or $3N$ normal vibrations) and will be determined in this chapter.

10.2.2. *Oscillations within a linear chain of atoms*

10.2.2.1. *Form of the solutions*

Here we study the longitudinal vibrations in a chain of N identical atoms that each have a mass M and are at equilibrium distances a apart. The length of the chain is thus given by $L = (N - 1)a \simeq Na$, as N is very high ($N >> 1$).This example uses a simple cubic structure and the propagation in the directions [100], [110] and [111] are considered. This means that the vibration goes along a chain of atoms.

The frequency of the oscillatory movement of one atom will be a function of the wave number K (3D vector) of the supposedly elastic vibration of the lattice.

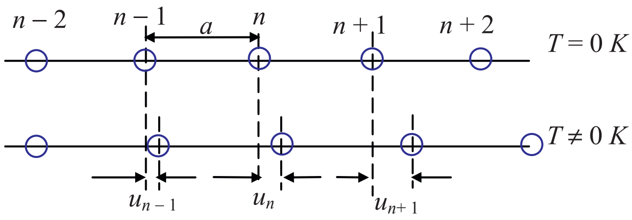


Figure 10.1. *Displacement of atoms at T ≠ 0 K through thermal agitation*

The spread of an atom denoted n from its equilibrium position by thermal agitation is denoted by u_n. Each atom is tied to its neighbors by a bond and, following the movements of the atoms, there is at the level of each atom a set of forces that tends to correct the movements. To simplify the problem, we can assume that each atom can only respond to the influence of its two nearest neighbors (any coupling beyond this is neglected), so that the atom numbered n is subject to only two forces, i.e. a force due to a bond with the atom numbered n - 1 and the other with $n + 1$. Assuming a linear approximation, these two forces are each proportional to the variation in the distance separating the atoms. If β denotes the proportionality constant (which shows the elastic properties of the lattice), the force against an atom n is given by:

$$F_n = -\beta\left(u_n - u_{n-1}\right) - \beta\left(u_n - u_{n+1}\right)$$

and the fundamental dynamic equation for this atom labeled n is:

$$M\frac{d^2u_n}{dt^2} = \beta\left(u_{n+1} + u_{n-1} - 2u_n\right) \qquad [10.1]$$

We can now look for solutions in the form of progressive plane waves of the type:

$$u(r,t) = A\exp\left[i\left(Kr - \omega t\right)\right]$$

where $K = \frac{2\pi}{\lambda}$ is the wave number and $\omega = 2\pi\nu$ is the angular frequency (also known as pulsation).

This means looking for the solution for the movement of the atom n located by $r = na$, and such that $u_n(na,t) = A\exp\left[i\left(Kna - \omega t\right)\right]$. Similarly, we also look for:

$$u_{n+1} = A\exp\left[i\left(K\left(n+1\right)a - \omega t\right)\right]$$

and:

$$u_{n-1} = A\exp\left[i\left(K\left(n-1\right)a - \omega t\right)\right].$$

Substituting this into equation [10.1] we obtain:

$$-M\omega^2 = \beta\left[e^{iKa} + e^{-iKa} - 2\right] = 2\beta\left(\cos Ka - 1\right),$$

so that, using $2\sin^2 x = 1 - \cos 2x$, we find:

$$\omega = 2\sqrt{\frac{\beta}{M}}\left|\sin\frac{Ka}{2}\right| \tag{10.2}$$

This is a dispersion equation based on $\omega = f(K)$, and is periodic with respect to K with a period given by $\frac{2\pi}{a}$ as:

$$\omega(K) = 2\sqrt{\frac{\beta}{M}}\left|\sin\frac{Ka}{2}\right| = \omega(K + \frac{2\pi}{a}) = 2\sqrt{\frac{\beta}{M}}\left|\sin\left(\frac{Ka}{2} + \pi\right)\right|$$

The representation is given in Figure 10.2a. However, given the periodicity it suffices in fact to represent the function ω(K) in an interval equal to $\frac{2\pi}{a}$ which is chosen to be between $-\frac{\pi}{a}$ and $\frac{\pi}{a}$ (Figure 10.2b). The width of the first Brillouin zone contains non-equivalent K values (the positive and negative values represent the two opposite directions in which the wave can move). This can be compared with the results in the following section.

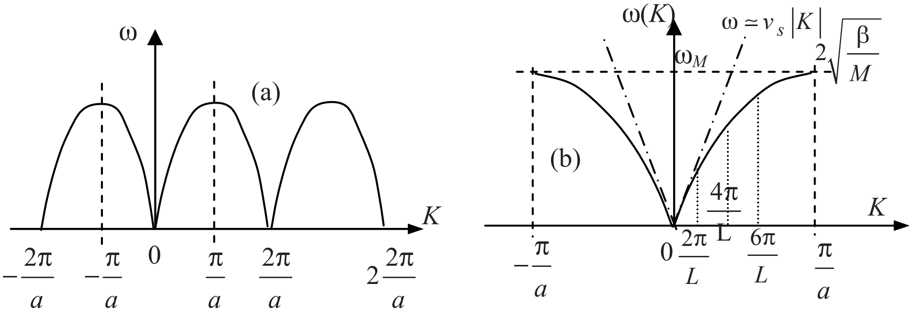


Figure 10.2. *(a) Dispersion curve [ω(K)] for a chain of atoms; and (b) representation for the first Brillouin zone*

10.2.2.2. *System properties*

– *Variations in the useful K zone.* We have $\frac{u_{n+1}}{u_n} = e^{iKa} = e^{i\varphi}$. When $Ka = \varphi$ between $-\pi$ and $+\pi$, Ka takes n all possible independent values that give the dephasing (φ) in the movement of two neighboring atoms. For example, if two atoms are out of phase by $\varphi = \left[\pi + \frac{x}{10}\pi\right]$ (with $0 < x \leq 10$) it is the same as stating that they are dephased by $\varphi = \left[\frac{10-x}{10}\right]\pi$. The range of variation in K is thus $-\frac{\pi}{a} \leq K \leq \frac{\pi}{a}$.

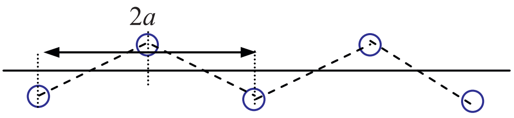
– *Difference with the continuous medium (where a → 0), and number of modes for a long chain.* For the difference in a continuous medium where a → 0, such that $K \to \pm\infty$ (as $K_{\max} = \pm\pi/a$), with the example of a discrete chain of atoms, the extreme values of K are $K_{\max} = \pm\frac{\pi}{a}$. For a progressive wave solution, such that in terms of r we would have $u(r,t) = A\exp\left[i\left(Kr - \omega t\right)\right]$, the use of the condition of periodic limits $u(r) = u(r + L)$ results in (see also section 2.4) $K = K_n = n\frac{2\pi}{L}$, where $n = 0, \pm1, \pm2....$ Here the quantification is located at n as N represents the number of atoms in the chain such that $L = Na$. These values are the permitted values for K in the expression for the progressive wave $u(r,t)$ solution for the system. The spectrum given by $\omega = f(K)$ is therefore discrete (see Figure 10.2b) and the interval denoted ΔK between two successive K values is simply $\Delta K = \frac{2\pi}{L}$, such that with the width of the Brillouin zone being given by $\Delta K_{Bril} = \frac{2\pi}{a}$, K can take on $\frac{2\pi/a}{2\pi/L} = \frac{L}{a} = N$ values.

– *Number of modes for a system with a limited extension (length L = Na where is small and will give* preference *to stationary systems – see also section 2.3).* In this case, there is a node at both extremities of the chain and the limiting conditions (at r = 0 and r = L where we should have $u(0,t) = u(L,t) = 0$), results in a stationary solution of the form (see section 2.3): $u(r) = A \sin Kr$, where $K = K_n = \frac{\pi}{L} n$ and n = 1, 2, 3... and only takes on positive values (n = 0 would give $u(r)$ = 0, so that there would be no more vibrations). Here $\Delta K = \frac{\pi}{L}$, and the spectrum ω = $f(K)$ is also discrete and retains the same shape with intervals between 0 and π/a, to which the system is limited in this configuration. For this stationary system and over the with π/a (zone of presence as K > 0), the number of normal modes possible is therefore given by $\frac{\pi/a}{\pi/L} = \frac{L}{a} = N$. This is the number of normal vibrational modes for this stationary system (identical to the number of modes obtained in the preceding section for a long chain).

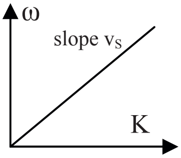
– *Mode densities.* With the interval $\Delta K = \frac{\pi}{L}$ corresponding to a vibrational mode, such that over a unit interval of K space we can "place" $\frac{1}{\pi/L} = \frac{L}{\pi}$ modes. This is the mode density in the K space. The mode density [D(ω)] in the pulsation space must therefore be such that $D(\omega)\, d\omega = \frac{L}{\pi} dK$, an expression that can also be written as $D(\omega)\, d\omega = \frac{L}{\pi} \frac{d\omega}{d\omega/dK}$. The velocity group ($d$ω/$dK$) appears in the denominator and this can be obtained from the dispersion curve. If the group velocity is zero (tangent horizontal to the dispersion curve), D(ω) appears as a singularity.

– *Crystal vibration frequency.* Equation [10.2] shows that the crystal can vibrate between ω = 0 and $\omega_M = 2\sqrt{\frac{\beta}{M}}$ (see Figure 10.2b).

– *The shortest and longest wavelengths.* As $\lambda_n = \frac{2\pi}{K_n} = \frac{2L}{n}$, we have $\lambda_{\min} \overset{n=N}{=} 2a$ while $\lambda_{\max} \overset{n=1}{=} 2L$. The shortest wavelength results directly from the discontinuous chain of the atomic chain for which the concept of a wavelength with λ < 2a, i.e. one that is associated with at least three atoms, would not make any sense.



– *Propagation velocity and wave frequency.* For a small value of K (equivalent to having large wavelengths corresponding to acoustics vibrational modes), an approximation can be made as in $\sin\frac{Ka}{2} \simeq \frac{Ka}{2}$. Equation [10.2] then becomes linear, as in $\omega = a\sqrt{\frac{\beta}{M}}\left|K\right| = v_s\left|K\right|$ and we have the straight line $\omega = v_s\,|K|$ from Figure 10.2b. We thus find a remarkable property in that $v_\varphi = \frac{\omega}{K} = v_s = \frac{d\omega}{dK} = v_g$. Thus the phase velocity and the group velocity have the same value, both equal to v_s or rather the velocity of sound in solids (of the order of 10^5 cm s^{-1}). This can be compared with the figure above and section 7.2.1 of [MOL 07b].



Thus setting $v_s = a\sqrt{\beta/M}$, expression [10.2] can be written as:

$$\omega = 2\frac{v_s}{a}\left|\sin\frac{Ka}{2}\right| \qquad [10.3]$$

For $k = \frac{\pi}{a}$, we have:

$$v_g = \left(\frac{d\omega}{dk}\right)_{k=\frac{\pi}{a}} = 2\sqrt{\frac{\beta}{M}}\frac{a}{2}\left|\cos\frac{ka}{2}\right|_{\frac{ka}{2}=\frac{\pi}{2}} = 0$$

The dispersion curve gives rise to a horizontal tangent with respect to $k = \pm\frac{\pi}{a}$ (see Figure 10.2b). The wave at this point is stationary with a maximum angular frequency (pulsation) (see Figure 10.2) with a value deduced from equation [10.3] equal to $\omega = \omega_{\max} = 2\frac{v_s}{a}$. With $v_s \approx 10^5$ cm s^{-1} and a ≈ 10^{-8} cm, we find that $\nu_{\max} = \omega_{\max}/2\pi \approx 10^{13}$ Hz, which is in the infrared region.

10.2.3. *Oscillations within a diatomic and 1D chain*

This section considers a chain made up of two types of atoms with masses m and M ($M > m$), each regularly spaced by a distance denoted a. The atoms with mass m are assumed to be placed at every $2na$ and those of mass M are at $(2na \pm a) = (2n \pm 1)a$, as shown in Figure 10.3.

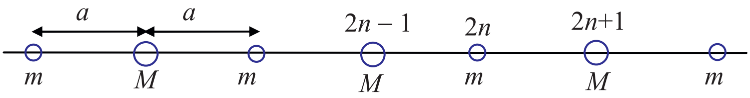


Figure 10.3. *Diatomic chain*

Supposing that each atom interacts, once again, only with its closest two neighbors, and that the interaction constants (β) are identical between paired neighbors, the equations for movement can be written as:

$$\begin{cases} -\text{ for atoms of mass } m\text{, then } m\dfrac{d^2u_{2n}}{dt^2} = \beta\left(u_{2n+1} + u_{2n-1} - 2u_{2n}\right) \\ -\text{ for atoms of mass } M\text{, then } M\dfrac{d^2u_{2n+1}}{dt^2} = \beta\left(u_{2n+2} + u_{2n} - 2u_{2n+1}\right) \end{cases} \qquad [10.4]$$

Once again we are looking for solutions in the form of progressive waves, but with amplitudes and frequencies different for each type of atom (as their mass is different):

$$\begin{cases} u_{2n} = A\exp\left[j\left(\omega_1 t - 2Kna\right)\right] \\ u_{2n+1} = B\exp\left[j\left(\omega_2 t - [2n+1]Ka\right)\right] \end{cases} \qquad [10.5]$$

Given the form of these solution, we thus have:

$$u_{2n+2} = u_{2n}\exp[-2jKa] \text{ and } u_{2n-1} = u_{2n+1}\exp[2jKa]$$

On introducing [10.5] into [10.4] we find that:

$$\begin{cases} -m\omega_1^2 u_{2n} = \beta\left[\left(1+\exp\left[2jKa\right]\right)u_{2n+1} - 2u_{2n}\right] & [10.6] \\ -m\omega_2^2 u_{2n+1} = \beta\left[\left(1+\exp\left[-2jKa\right]\right)u_{2n} - 2u_{2n+1}\right] & [10.7] \end{cases}$$

From equation [10.7] we can pull out that:

$$u_{2n+1} = \frac{\beta\left[1+\exp\left(-2jKa\right)\right]}{2\beta - M\omega_2^2} u_{2n} \qquad [10.8]$$

Equation [10.8] must be true for all values of time (*t*). According to [10.5], u_{2n} varies with ω_1 and u_{2n+1} as a function of ω_2. We must therefore find that $\omega_1 = \omega_2 = \omega$.

Inserting equation [10.8] into [10.6], we find that:

$$\left(2\beta - M\omega^2\right)\left(2\beta - m\omega^2\right) - 4\beta^2\cos^2 Ka = 0$$

From this we deduce that:

$$\omega^4 - 2\beta\frac{M+m}{Mm}\omega^2 + 4\frac{\beta^2\sin^2 Ka}{mM} = 0 \qquad [10.9]$$

This equation has two solutions for ω^2 which can be denoted ω_+^2 and ω_-^2 and are such that:

$$\omega_\pm^2 = \beta\frac{m+M}{mM}\left[1 \pm \sqrt{\frac{4mM\sin^2 Ka}{\left(m+M\right)^2}}\right] \qquad [10.10]$$

so that also:

$$\omega_\pm^2 = \beta\left(\frac{1}{m}+\frac{1}{M}\right) \pm \beta\sqrt{\left(\frac{1}{m}+\frac{1}{M}\right)^2 - \frac{4\sin^2 Ka}{mM}} \qquad [10.11]$$

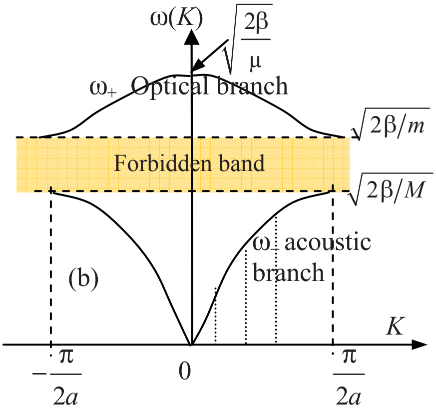


Figure 10.4. *Acoustic and optical branches, with M > m*

This result (equation [10.11]) shows that for a chain made up of atoms of varying nature there can be two types of vibrations with different frequencies, namely ω_+ and ω_-. In other terms, a value for K can correspond to two values for ω, and therefore two vibrational modes.

The solution ω_+ (upper branch of Figure 10.4) shows the optical branch of the vibrational spectrum, whereas the ω_- solution (lower branch of Figure 10.4) is associated with the acoustic domain (low frequencies).

NOTE 1.– Using $2\sin^2 Ka = 1-\cos 2Ka$, equation [10.11] can be rewritten as:

$$\omega_\pm^2 = \beta\left(\frac{1}{m}+\frac{1}{M}\right) \pm \beta\sqrt{\left(\frac{1}{m}+\frac{1}{M}\right)^2 - \frac{2\left(1-\cos 2Ka\right)}{mM}}$$

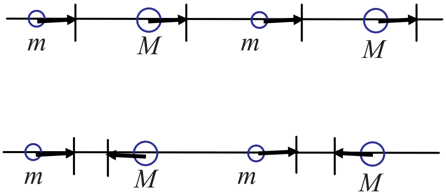
– When Ka is small we have $\cos 2Ka \simeq 1$, and $\omega_+^2 \simeq 2\beta\left(\frac{1}{m}+\frac{1}{M}\right)$, so that with $\frac{1}{\mu} = \frac{1}{m}+\frac{1}{M}$ (reduced mass) then $\omega_+ = \sqrt{\frac{2\beta}{\mu}}$.

Still in the condition of Ka being small, we have $\left(\omega_-\right)_{K\simeq 0} \simeq 0$.

– When $K = \pm\frac{\pi}{2a}$, such that $Ka = \pm\frac{\pi}{2}$ which gives $\cos 2Ka = \cos(\pm\pi) \;= -1$, we have:

$$\begin{aligned}\omega_{\pm}^2 &= \beta\left(\frac{1}{m}+\frac{1}{M}\right) \pm \beta\sqrt{\left(\frac{1}{m}+\frac{1}{M}\right)^2 - \frac{4}{mM}} \\ &= \beta\left(\frac{1}{m}+\frac{1}{M}\right) \pm \beta\sqrt{\left(\frac{1}{m}-\frac{1}{M}\right)^2}\end{aligned}$$

from which $\omega_+ = \sqrt{\frac{2\beta}{m}}$ and $\omega_- = \sqrt{\frac{2\beta}{M}}$.



NOTE 2.

– In the acoustic branch with modes for very low frequencies ($K \approx 0$ and $\omega_- \approx 0$), equation [10.8] shows that $u_{2n+1} \approx u_{2n}$. This means that the atoms and the centers of their mass vibrate together (identical direction) and there is a translation of the complete cell, as shown in the top half of the figure above.

– For the optical branch, and when $K \approx 0$, using $\omega_+^2 \simeq 2\beta\left(\frac{1}{m}+\frac{1}{M}\right)$ in equation [10.8] we find that:

$$\frac{u_{2n+1}}{u_{2n}} = \frac{2\beta}{2\beta - M2\beta\left(\frac{1}{m}+\frac{1}{M}\right)} = -\frac{m}{M}$$

and the atoms vibrate with respect to one another but with the center of their mass fixed, as shown in the bottom part of the figure above.

NOTE 3.

– If two atoms carry opposing charges, we can imagine generate a vibration of this type using the electric field of a light wave – hence the name optical branch.

– Vibrations in the acoustic branch are tied to thermal effects and the passing of sound.

10.2.4. *Vibrations of a 3D crystal*

– *Transverse and longitudinal modes.* In 3D, the wave vector becomes 3D and the vector gives three corresponding vibrational modes. In addition to the longitudinal mode of the preceding section, there are also two transverse modes polarized at 90° to each other (for a definition of the various mode types see, for example, section 6.2 of [MOL 07b]). These two new transverse modes are also characterized by the curve ω(K) carrying acoustic and optical branches and a forbidden band.

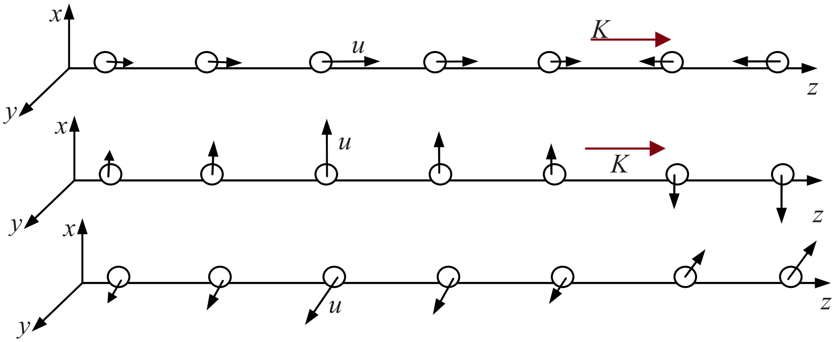


Figure 10.5. *Vibrational modes for a chain of atoms in a 3D lattice: (a) longitudinal modes; and (b) transverse mode*

– *Number of modes.* If the whole crystal encloses N atoms, then we have in all $3N$ vibrational modes with $2N$ of those being due to transverse vibrations and the other N being longitudinal (see problem 10.6.1).

– If the lattice contains two types of atoms (section 10.2.3), and taking into account note 2 of the preceding section, for the transverse modes and along the acoustic branch the two types of atoms oscillate in phase. However, along the optical branch, they vibrate in a mutually opposing phase.

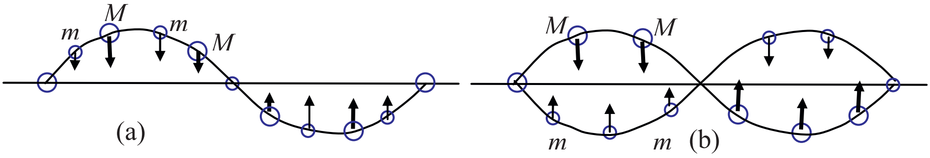


Figure 10.6. *Transverse modes for a lattice made up of two types of atoms: (a) acoustic branch; and (b) optical branch*

10.2.5. *Energy of a vibrational mode*

Here we use Debye's theory where it is supposed that:

– Atomic oscillations in a lattice resemble a harmonic oscillator. The $3N$ vibrational modes are assumed to correspond to the same number of harmonic oscillators which have energy levels that are quantified by:

$$E = E_n = \left(n + \frac{1}{2}\right)\hbar\omega_0 , \qquad n = 0, 1, 2...$$

(see a course on quantum mechanics). The proper frequencies of the oscillator are equally spaced by an interval of ω_0. In Debye's model, we can simplify the relatively complicated dispersion curve (and given by equations [10.2] or [10.11] for example) by considering the right-hand side, $\omega = v_s|K|$, no longer with the terms $K_{\max} = \frac{\pi}{a}$, but rather with just K_{m} being the total number of unnumbered modes equal to $3N$. Thus the K_n values of K are regularly spaced, the dispersion curve straightened, and the corresponding values for ω_n are also regularly separated by $\Delta\omega \overset{notation}{=} \omega \overset{Debye}{=} \omega_0$, just as in a harmonic oscillator.

– Oscillators are energetically distributed according to Boltzmann statistics. This indicates that the probability that a particular mode will have an energy E_n is proportional to $\exp\left(-\frac{E_n}{kT}\right)$.

So, at thermal equilibrium, the average energy for a pulsation mode given by $\omega_0 = \omega$ is written as:

$$\overline{E_\omega} = \frac{\sum\limits_{n=0}^{\infty}\left(n+\frac{1}{2}\right)\hbar\omega \exp\left[-\left(n+\frac{1}{2}\right)\hbar\omega \big/ kT\right]}{\sum\limits_{n=0}^{\infty}\exp\left[-\left(n+\frac{1}{2}\right)\hbar\omega \big/ kT\right]} \qquad [10.12]$$

By making $x=-\frac{\hbar\omega}{kT}$, we obtain:

$$\begin{aligned}\overline{E_\omega}&=\hbar\omega\frac{\frac{1}{2}e^{x/2}+\frac{3}{2}e^{3x/2}+\ldots}{e^{x/2}+e^{3x/2}+\ldots}\\&=\hbar\omega\frac{d}{dx}\left\{\mathrm{Ln}\left[e^{x/2}\left(1+e^{x}+e^{2x}+\ldots\right)\right]\right\}\\&=\hbar\omega\frac{d}{dx}\left\{\mathrm{Ln}\left[e^{x/2}\frac{1}{1-e^{x}}\right]\right\}\\&=\hbar\omega\frac{d}{dx}\left[\frac{1}{2}x-\mathrm{Ln}\left(1-e^{x}\right)\right]=\hbar\omega\left[\frac{1}{2}+\frac{1}{\exp\left(\hbar\omega/kT\right)-1}\right]\end{aligned}\qquad[10.13]$$

If $T = 0K$, then $\left(\overline{E_\omega}\right)_{T=0K}=\frac{1}{2}\hbar\omega$ and this is the energy of what are generally called the lattice zero vibrations. They are normally ignored when studying the thermal properties of crystals. In this approximation, we can therefore write from equation [10.13] that:

$$\overline{E_\omega}\simeq\frac{\hbar\omega}{\exp\left(\hbar\omega/kT\right)-1}\qquad[10.14]$$

NOTE.– The vibrational modes where $\hbar\omega << kT$, with $kT \approx 0.026$ eV at $T = 300\ K$ (i.e. acoustic modes with low energies and long wavelengths) have, according to equation [10.14], an average energy given by:

$$\overline{E}_{\text{acoustic}}=\overline{E}_A\simeq\frac{\hbar\omega}{1+\frac{\hbar\omega}{kT}-1}=kT.\qquad[10.15]$$

When these conditions are present, the mode is termed as being completely excited. If T is sufficiently high so that this can be spread to $3N$ modes, the internal vibrational energy of the crystal is equal to $3NkT$.

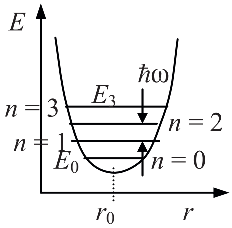


Figure 10.7a. *Levels in a harmonic oscillator*

10.2.6. *Phonons*

10.2.6.1. *Definition*

Charge carriers (electrons, holes) continuously interact with the lattice, and give or gain energy from it. These exchanges of energy operate through transitions in vibrational modes. As the energy of each vibrational mode is quantified, the exchanges can only be quantified between energy levels of the form $E_n = (n+1/2)\hbar\omega$. The transitions are in accordance with the selection rule $\Delta n = \pm 1$, and the carrier–lattice interactions occur through absorption or emission (generation) of $\hbar\omega$ quanta of energy.

By analogy with photons, these quanta are termed phonons and they too can lead a double life by exhibiting both wave and particulate natures. They are characterized by having:

– energy $E_{pn} = \hbar\omega$

– quantity of movement $\hbar K$

– zero mass, like a photon

– and an absence of spin – such that the statistical form used to study them is that of Bose-Einstein (see equation [10.14] in which there is the Bose-Einstein occupation factor given as $\left[\exp(\hbar\omega / kT - 1)\right]^{-1}$).

10.2.6.2. *Properties of phonons and the physical role of semiconductors*

Each normal pulsation (ω) contains *n* phonons of energy $\hbar\omega$. This energy is, in reality, a pretty small amount. As we saw above, even when fully excited (in the acoustic mode), then $\hbar\omega << kT \approx 0.026$ eV. However, the amount of movement can actually be quite high, as *K* can reach $\frac{\pi}{a}$ which is of the order of the quantity of movement of electrons in the Brillouin zone.

The electron-lattice vibration interaction can therefore be thought of as the interaction of two bodies, namely electrons and phonons with a conservation of energy and a quantity of movement.

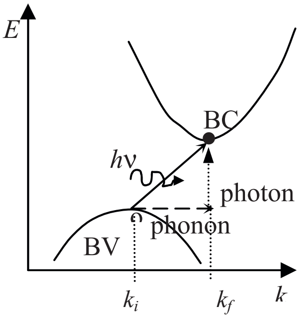


Figure 10.7b. *Indirect (oblique) transition causing introduction of a phonon*

Given a phonon's characteristics, a collision between an electron and a phonon can little change the energy of the electron but may greatly change its quantity of movement. This is the mechanism used in transitions between permitted bands in indirect gap semiconductors (from the extrema of valence and conduction bands obtained for wave vector different values, i.e. k_i and k_f as shown in Figure 10.7b). The variation in energy is assured by the photon that produces the vertical transition (i.e. without change in quantity of movement) and the phonon that gives the quantity of movement, as in $\hbar\left(k_i - k_f\right)$ which makes an oblique transition possible. Numerically, the wave vector for the photon is given by $k_{pt} = \frac{2\pi}{\lambda} \overset{\lambda \approx 10^3 nm}{\approx} 10^{-2}\, nm^{-1}$, while that of the phonon is better written as $K_{pn} \approx \frac{\pi}{a} \overset{a \simeq 0.3nm}{\approx} 10\; nm^{-1}$.

10.2.7. *Conclusion*

To resume, it is possible to represent, indifferently, the vibratory states of a crystal:

– either as a sum of $3N$ vibrational modes of which the average energy for a pulsation mode (otherwise known as the angular frequency and denoted ω) is given by $\overline{E_\omega} \simeq \frac{\hbar\omega}{\exp\left(\hbar\omega/kT\right)-1}$;

– or as a free "phonon gas" in which the quasi-particles exhibit their double character of wave or particle with a low energy and a high degree of movement.

It is also possible to state that, in contrast to electrons, the number of phonons for a given energy state is not limited. There is a constant annihilation and creation of phonons in the lattice. The conservation of the number of phonons present is not obligatory.

10.3. Polarons

10.3.1. *Introduction: definition and origin*

Simplistically put, we can think of electrons or holes moving in a crystal. Due to the charges they carry, they displace the ions, which are of opposite charge, in the lattice and can even create a polarization associated with the resulting deformation. This force for such a change is occasionally called a constraining field. The charged particle (electron or hole) and the associated deformation (by polarization) form the so-called polaron quasi-particle. As the charged particle moves in the crystal, it “pulls” on the deformation increases its effective mass and correlatively decrease its mobility (see section 8.1.3).

The polaron can be thought of as an electron-lattice coupling, as it is the latter that provokes, in some senses at least, the localization of the charge into the deformation where it represents a reduced mobility with respect to a configuration where the deformation of the lattice would be inexistent or negligible. The latter case happens in covalent crystals as these are made up of neutral atoms which have weak interactions with the charges (electrons or holes). The effects are considerably greater in ionic crystals because of the strong Coulombic interaction between charges and ions.

The pairing can also be thought of as an electron-phonon coupling, where it is the longitudinal phonons that stimulate deformations and with them propagate with the charges. This is a mechanism that would explain the transport of polarons. In the case of non-polar solids, the dominant interaction is with acoustic longitudinal polarons. However, in polar media, it is the optical longitudinal polarons that are strongly coupled with electrons. This is due to great variations in the active dipolar moments associated with atomic displacements in opposition to the phases induced by optical phonon modes (see also note 2 in section 10.2.3).

10.3.2. *The various polarons*

10.3.2.1. *Dielectric polarons*

If we take an ionic crystal lattice, as shown in Figure 10.8a, and then place an electron on an ion, as detailed by the black point in Figure 10.8b, then we can see that surrounding ions undergo a force due to the additional electron. This electron–lattice interaction results in new positions for the ions, as shown in the same figure by dotted lines. This displacement of ions always results in a reduction of the energy of the electron, and also results in a potential well within which can be found the electron. If the well is deep enough then the electron will find itself in a tied state, incapable of moving to another site unless there is modification in the positions of neighboring ions. We can label the electron "self-trapped" and as such, it and its associated lattice deformation is termed a "polaron" The term originated from phenomena observed in polar materials; however, such quasi-particles can also occur in covalent materials.

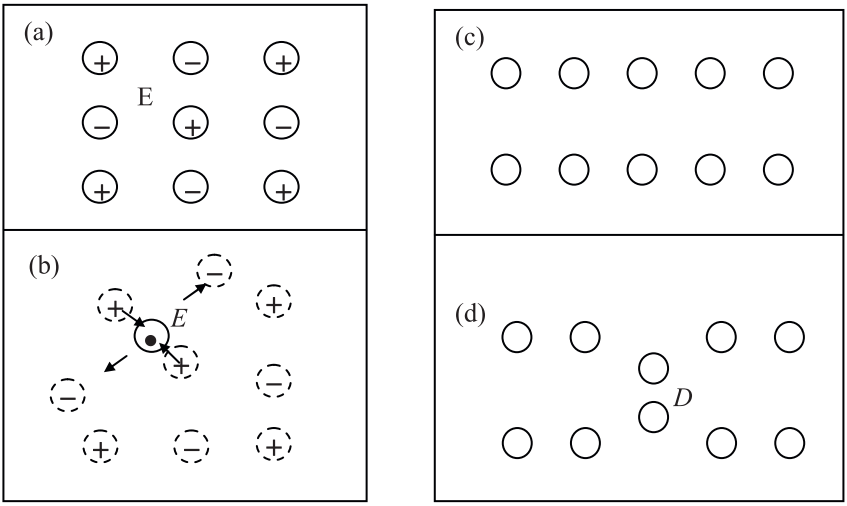


Figure 10.8. *(a) Lattice of ions in relaxed state; (b) repositioning of ions in directions of arrows following placement of electron at E in the lattice; (c) and (d) polaron formation in covalent materials going from a regular arrangement of rare gas atoms in (c) to a deformed atomic lattice after a hole has been placed at D.*

10.3.2.2. *Molecular polarons*

Molecular polarons form in covalent materials in which the resulting distortion is confined to neighboring atoms, which can subsequently form a chemical bond while the charge is trapped. A good example is of V_k centers in alkali metal-halogen crystals, in which a hole trapped on a Cl^- ion results in attracting a neighboring Cl^- ion to yield a "molecular ion" of form Cl_2^-. Similar phenomena can occur in solid rare gases in which there is a trapped hole, as detailed in Figures 10.8c and d, and similarly in certain mineral glasses in which dangling bonds at a neutral site (D^0) can result in a more favorable local rearrangement such that $2\ D^0 \rightarrow D^+ + D^-$ where D^+

and D^- are the previously neutral dangling bonds (D^0) that have, respectively, lost or gained an electronic charge.

Section 10.3.4 will detail the origin of polarons in molecular crystals (using the Holstein model).

10.3.2.3. *Small and large polarons*

If the wavefunction associated with a self-trapped electron takes up a space equal to or smaller than the lattice constant, then the polaron is called a "small polaron" and the deformation is localized only in the neighborhood of the charge carrier. This type of quasi-particle is formed in covalent materials as they are essentially made up of neutral atoms which undergo only weak interactions with the charges (electrons or holes). The dielectric polaron can itself result in a small polaron when the distortion of the lattice is limited to the immediate neighborhood of the charge (as in Figure 10.8b). In the opposing case, then a large polaron, otherwise termed a "Fröhlich polaron", can be formed in polar media in which Coulombic forces are involved that polarize the crystal over long distances (such as in metallic oxides). Small and large dielectric polarons are distinguished by a determination of the Fröhlich polaron coupling constant.

10.3.3. *Dielectric polarons*

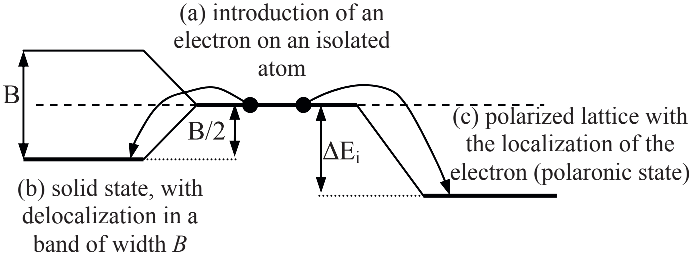


Figure 10.9. *Energy levels for an electron introduced in: (a) an isolated atom; (b) a solid wherein it delocalizes; and (c) in a polar lattice where is becomes localized*

If an electron is introduced into an unfilled orbital of a given atom in a solid, this electron can display one of two types of behavior, depending on the nature of the solid. They are:

– either the electron delocalizes into the bands resulting from an overlap of the orbitals of the atoms in the solid, and the electron is stabilized in the solid due to a decrease in its energy by an amount that is of the order of $B/2$, where B represents the width of the permitted band (as in Figure 10.9a);

– or the electron will polarize the crystal, if the crystal will allow it, and there will be a displacement of neighboring atoms resulting in a deformation by polarization of the lattice. If it is assumed that the each electronic charge, denoted e, is spread around a spherical orbital with a radius r (associated with the assumed deformation of the lattice) then the energy of the charged sphere (assumed to be in equilibrium) can be written as $E_p = \frac{1}{2}\frac{e^2}{C}$. Here, C is the capacitance of the sphere and is given by $C = 4\pi\varepsilon_0\varepsilon_r r$ where ε_r is the relative permittivity. The change in the energy of the polarization by the charge in going from vacuum (localized on a single atom surrounded by a vacuum) to solid states, where the solid has a relative permittivity ε_r, is given by:

$$\begin{aligned}\Delta E_{po} &= E_{\substack{\text{solid}\\ \text{lattice}}} - E_{\text{vacuum}} = \left(\frac{1}{2}\frac{e^2}{4\pi r\varepsilon_0\varepsilon_r r}\right) - \left(\frac{1}{2}\frac{e^2}{4\pi r\varepsilon_0 r}\right)\\ &= -\frac{e^2}{8\pi r\varepsilon_0 r}\left[1-\frac{1}{\varepsilon_r}\right]\end{aligned} \qquad [10.16]$$

If we assume that the lattice deformation is only produced by the ionic and electronic polarizations, then the permittivity can be limited to the ionic and electronic components, as in $\varepsilon_r = \varepsilon_{\text{ion}} + \varepsilon_{\text{opt}}$, where ε_{opt}, the electronic component only appears in the optical domain (see section 3.4 of [MOL 07a]). ε_r is the resulting permittivity which accounts for the establishment of the various polarizations (which *a priori* takes an infinite time) or zero frequency if ε_r is static permittivity.

The decrease in electronic polarization energy is thus given by substituting ε_{opt} for ε_r in [10.16], as in:

$$\Delta E_e = -\frac{e^2}{8\pi r\varepsilon_0 r}\left[1-\frac{1}{\varepsilon_{opt}}\right] \qquad [10.17]$$

With $\Delta E_{po} = \Delta E_e + \Delta E_i$, the decrease in energy (ΔE_i) by the single ionic polarization is given by:

$$\begin{aligned}\Delta E_i &= -\frac{e^2}{8\pi r\varepsilon_0 r}\left[1-\frac{1}{\varepsilon_r}\right]-\left(-\frac{e^2}{8\pi r\varepsilon_0 r}\left[1-\frac{1}{\varepsilon_{opt}}\right]\right)\\ &= -\frac{e^2}{8\pi r\varepsilon_0 r}\left[\frac{1}{\varepsilon_{opt}}-\frac{1}{\varepsilon_r}\right]\end{aligned} \qquad [10.18]$$

Returning to Figure 10.8, it is possible to see that if $\Delta E_i > B/2$, the localized state (polaronic) is more stable than that that the electron would use. This condition can be seen as a proviso for the formation of a polaron. In this polaronic state, where the electron is spread over a radius r, its kinetic energy resembles that which it would have were it in a potential sphere of radius r such that

$$E_{\text{kin}} = \frac{\hbar^2 k^2}{2m^*} \approx \frac{h^2}{2m^*[2\pi]^2}\left(\frac{[2\pi]^2}{r^2}\right) \approx \frac{h^2}{2m^* r^2}$$

(the effective mass accounts for the fact that the electron is tied to the lattice and is therefore not free).

The total energy of an electron localized in the sphere is thus the sum of this kinetic energy and the energy gained through lattice deformation (ionic polarization). By deriving the energy with respect to r, and by making it equal to zero, we can work out the value of r that gives a stable state:

$$r = \frac{8\pi\varepsilon_0 h^2}{m^* e^2\left(1/\varepsilon_{\text{opt}} - 1/\varepsilon_r\right)} \qquad [10.19]$$

A small value of r is favored by a large effective mass and a strong ionic polarization (that makes $[\varepsilon_r - \varepsilon_{\text{opt}}]$ large). In this case, and especially when m^* is large, with have a narrow band system with a low mobility. Finally, if $r < a$ (the lattice repeat unit) it can be said that we have a small polaron, however, if $r > a$, then it is called a large polaron.

10.3.4. *Polarons in molecular crystals*

10.3.4.1. *Holstein's model in equations*

Molecular crystals can be used to obtain a simple, generalized model for small polarons (as with Holstein's model in "Studies of polaron motion: Part I. The molecular-crystal model", *Annals of Physics,* 1959, 8, 325). Just one excess electronis placed in a regularly aligned, flexible lattice of molecules (mass M). With each molecule denoted g we associate a coordinate x_g that represents the movement of the molecule under a harmonic vibration of pulsation ω_0 with frequency:

$$\nu_0 = \frac{1}{2\pi}\sqrt{\frac{k}{M}}$$

Starting with a single molecular site and excluding coupling effects with surrounding molecules, $\zeta(r)$ represents the potential energy of an oscillator and the vibrational force is given by $\vec{F}(r) = -\overrightarrow{\text{grad}}\,\zeta(r)$. When r_0 is in an equilibrium position, such that $x_g = \Delta r = r - r_0$ represents the local deformation of a molecule, we have (see section 10.2.1),

$$W(r) = W(r_0) + \frac{1}{2}\,kx_g^2\,,$$

where k is in accordance with the harmonic equation $F(r) = M\frac{d^2x_g}{dt^2} = -\,kx_g$. The solution, $x_g = X\cos\omega_0 t$, necessitates the introduction of an actual pulsation $\omega_0 = \sqrt{\frac{k}{M}}$, and finally we obtain (with $k = M\,\omega_0^2$):

$$W(r) = W(r_0) + \frac{1}{2}M\omega_0^2x_g^2$$

Then ω_0 can be estimated for an elongation (x_g) around about the same size as the lattice constant (a) because the vibrational energy given by $\frac{1}{2}M\omega_0^2a^2$ is of the same order as that of a bond energy (E_L) in the molecule and $E_L \approx 1$ eV. With $a \approx 1$ Å and $M = \frac{10^{-3}}{N}$ (hydrogen atom mass 10^{-3} kg), we have $\frac{1}{2}.\frac{10^{-3}}{6.02\text{x}10^{23}}(10^{-2})^2\ \omega_0^2 = 1.6 \times 10^{-19}$ J and thus $\omega_0 \approx 10^{14}$ rad s^{-1}. If a = 1 nm though, we now have $\omega_0 \approx 10^{13}$ rad s^{-1} and therefore, for heavier atoms, the frequencies are even lower.

On taking into account the coupling between the vibrational movement of a molecule and that of its neighbors, resulting in the transfer of vibrational energy throughout the lattice, we have to bring the phenomenon of frequency dispersion into play: $\omega = f(k)$. The Hamiltonian corresponding to the molecular crystal, without free charges, is written:

$$H_L = \sum_g \left\{ -\frac{\hbar^2}{2M}\frac{\partial^2}{\partial x_g^2} + \frac{M\omega_0^2 x_g^2}{2} + \sum_h M\omega_0\omega_b x_g x_{g+h} \right\}$$

where h is the nearest neighbor, M is the reduced mass and $6\omega_b$ is the size of the band of the optical phonons.

When an additional excess electron is introduced into the lattice, the electron-lattice interactions can be accounted for by considering the excess carrier energy at a site in the lattice. We accept that the energy is a linear function of movement within the lattice, and the greater the induced movement then the greater the absolute value of the electron-lattice energy and the more easily the charge is self-trapped. This trapping is actually greater than any coupling.

For a carrier localized on a site g, we can write $E_g = E_0 - \sum_{g'} f(g' - g)\, x_{g'}$, in which $f(g' - g)$ is a weighting factor which carries electron-lattice interactions. Assuming interactions are over short distances, $f(g' - g) = A\ \delta_{g'\,g}$, and the above equation changes to $E_g = E_0 - A\ x_g$, where E_0 is the energy of a non-distorted site, E_g is $E_0 = 0$ plus a constant and A represents the electron-site coupling force, as in $A = -\frac{E_g - E_0}{x_g - 0} = -\operatorname{grad} E$. In general, we write $E_g = E(x_g) = -Ax_g$.

Thus in the scenario we are considering, molecular deformations induced by the charge carrier are mostly localized around the carrier itself, and it is the presence of vibrational coupling between neighboring molecules that distribute distortion effects beyond the occupied site. Figure 10.10, otherwise known as an Emin representation, presents a scheme of the distance from equilibrium of diatomic molecules – represented by vertical lines – in a linear molecular crystal about a site with an electron (black dot).

The complete resolution of this problem will require the use of Hamiltonian operators (H_e) for strongly bonded electrons in a crystal of covalently bonded molecules, which will add a Hamiltonian for vibrational energies (H_L).

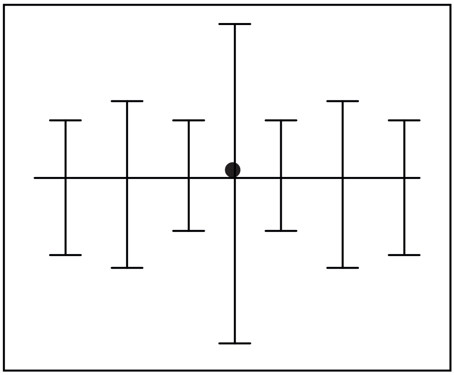


Figure 10.10. *Scheme of equilibrium separation distances for diatomic molecules in a linear chain, in which an electron (black dot) is placed at the center of a molecule*

10.3.4.2. *Limiting case of a single molecular site: polarons and bipolarons*

10.3.4.2.1. One electron

In this limited example, we will place one excess electron on a single site, the molecule of which undergoes a deformation x. The potential energy of the site can be expressed as:

$$E = \underbrace{M\ \omega_0^2\ \frac{x^2}{2}}_{\substack{\uparrow \\ \text{Vibrational} \\ \text{energy of} \\ \text{molecular} \\ \text{site}}} + \underbrace{(E_0 - A\ x)}_{\substack{\uparrow \\ \text{Energy of electron-} \\ \text{lattice interaction} \\ \text{(electron-vibration} \\ \text{coupling)}}} = B\ x^2 - A\ x \text{ where } E_0 = \underset{\substack{\uparrow \\ \text{Energy when} \\ \text{there is no} \\ \text{distortion}}}{0} \text{ and } B = \frac{1}{2} M\ \omega_0^2$$

E is optimized for a value $x = x_0$ such that:

$$\left(\frac{\partial \mathrm{E}}{\partial \mathrm{x}}\right)_{\mathrm{x}=\mathrm{x}_0} = 0, \text{ so that } M\,\omega_0^2\,x_0 - A = 2x_0\,B - A = 0 \Leftrightarrow x_0 = \frac{\mathrm{A}}{\mathrm{M}\omega_0^2} = \frac{\mathrm{A}}{2\mathrm{B}}$$

Following a deformation (x_0) of the lattice through vibration and electron-lattice interactions, the energy of electron-lattice interactions is accordingly reduced by Ax_0 ($-Ax_0$ written algebraically). Nevertheless, the system undergoes a distortion associated with a vibrational energy equal to $B\,x_0^2$ which is such that:

$$B\,x_0^{\,2} = B\left(\frac{A}{2B}\right)\,x_0 = \frac{1}{2}\,Ax_0.$$

Thus the energy of the system is decreased by E_p, where:

$$E_p = B\,x_0^{\,2} - A\,x_0 = -\,\frac{1}{2}\,Ax_0 = -\,\frac{\mathrm{A}}{4\mathrm{B}} = -\,\frac{\mathrm{A}^2}{2\mathrm{M}\omega_0^2}$$

The various energies are shown in Figure 10.11.

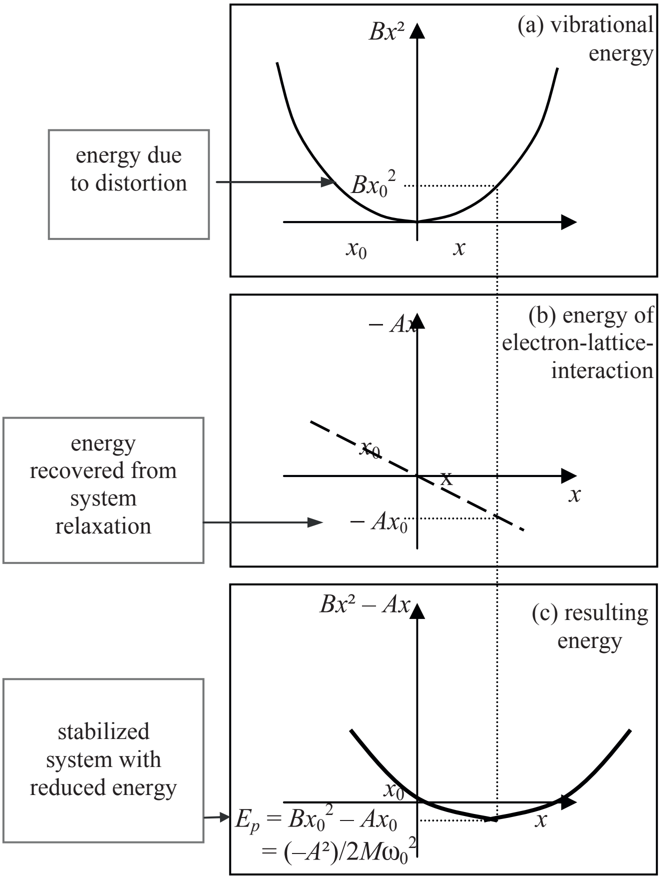


Figure 10.11. *Different energy terms with respect to lattice deformation by an excess electron*

Just as in polarized media, the formation of a polaron is favored by the distortion energy being less (in absolute terms exactly one half) than the energy recovered from the reduction in the electron–lattice interaction energy.

10.3.4.2.2. Two electrons: the bipolaron

It is possible to simply adapt the single site model to a situation where there are two electrons localized on a deformed molecule. The energy is now

$$E = \underset{\uparrow}{M\,\omega_0^2}\,\frac{x^2}{2} - \underset{\uparrow}{2A}\,x + U = B\,x^2 - (2A + C)\,x + U_0$$

virbrational energy

electron lattice coupling energy

Coulombic repulsion due to 2 electrons on the same site. We can suppose that U varies following $U = U_0 - Cx$, in which U_0 is the repulsion at zero deformation.

On replacing A in the preceding section with $2A + C$, the minimum E is now given by

$$x = x_1 = \frac{2A + C}{M\omega_0^2} \text{ and } E_{BP} = -\frac{(2A + C)^2}{2M\omega_0^2} + U_0$$

Two electrons localized on a single deformation is called a bipolaron and it is stable if the energy required for its formation is less than twice the energy of two isolated polarons, i.e. when

$$-\frac{(2A + C)^2}{2M\omega_0^2} + U_0 < -\frac{A^2}{M\omega_0^2}$$

The above relationship can be true when A and C are of the same sign and when U_0 is not too high. The deformation x_1 imposed by the two particles (here two electrons, but it could also be holes) can be advantageous in overcoming moderate Coulombic repulsions.

10.3.5. *Energy spectrum of the small polaron in molecular solids*

As we have just seen, an electron that distorts the molecule onto which it is placed reduces its energy $-Ax_0 = 2E_p$. The vibrational energy of the deformed molecule increases by $B\,x_0^2 = |E_p|$, and results in an overall reduction in energy for the system (equal to E_p) with respect to the energy of an electron in a rigid crystal of molecules where $x_g = 0$.

Within the limits of the preceding calculation, in which we have ignored vibrational dispersions due to coupling with non-existing neighboring molecules, it is possible to state that:

$$\left|E_p\right| = \frac{A^2}{2M\omega_0^2}$$

(see Figure 10.12).

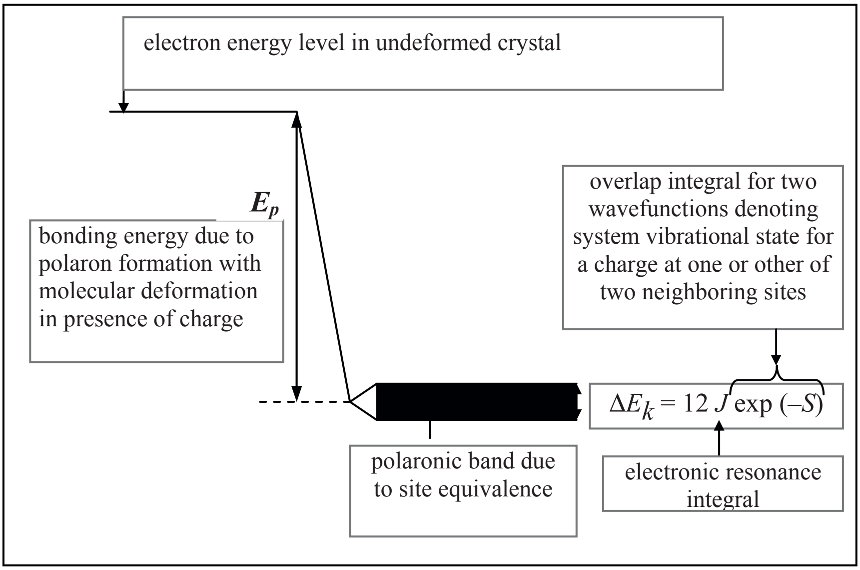


Figure 10.12. *Energy scheme for the small polaron*

If a small polaron (charge carrier and associated lattice deformation combined) can be equally situated at any other of the geometrical equivalents in a crystal, then we find that the actual states of the system are shared in a polaronic band as shown in Figure 10.12. Using a modified method of that used for strong bond approximations (equation [10.2] and the following equations from Chapter 8), we can consider that the proper states of a small polaron in a cubic crystal display an energy in the form:

$$E_k = -2\,J\exp(-\mathrm{S})\,[\cos k_x\,a + \cos k_y\,a + \cos k_z\,a] - E_p$$

where:

– $\mathrm{k} = \sqrt{\mathrm{k}_\mathrm{x}^2 + \mathrm{k}_\mathrm{y}^2 + \mathrm{k}_\mathrm{z}^2}$ is the polaron wave vector, and a the lattice constant;

– J is the resonance integral between two "electronically coupled" neighbors and is in the form $\exp(-\alpha R)$ to account for the exponential form of the electron wavefunctions; and

– $\exp(-S)$ is an overlap factor associated with chain vibrations. It represents the superposition integral between two wavefunctions, which detail the vibrational state of the system when the charge carrier is on one or another of two adjacent sites in a crystal. In the limiting case of a rigid lattice, the vibrational wavefunction remains unchanged during charge transfer from one site to its neighbor, and the overlap factor is equal to 1. However, here where we are looking at a distortable lattice, then $\exp(-S)$ can be considered to relate the necessary overlapping of atomic sites between which the required tunnel effect can occur to allow a complete displacement of atomic site and charge and is accounted for by J.

In fact, the transfer of a polaron requires two tunneling effects. One is associated with moving a charge between two neighboring sites (electron resonance integral), while the other is the movement of the deformation and any sites geometrically tied to the deformation. $\mathrm{Exp}(-S)$ is a factor of the same order as the atomic tunnelling effect as it assimilates atomic site transfers. As it is associated with the high mass of atoms, relative to the charge carriers, its value is extremely low. For a cubic crystal, the polaronic band size is $\Delta E_k = 12\,J\exp(-S)$ and is extremely narrow, of a width typically below that of vibrational energies, which are at least $kT \geq 10^{-4}$ eV at temperatures not too close to zero.

NOTE.– The displaced deformation is equivalent to an "exchange" in position for two neighboring sites as demonstrated in Figure 10.13. Evidently, it is not two actual atoms which change place but their position relative to the deformation propagated with a charge.

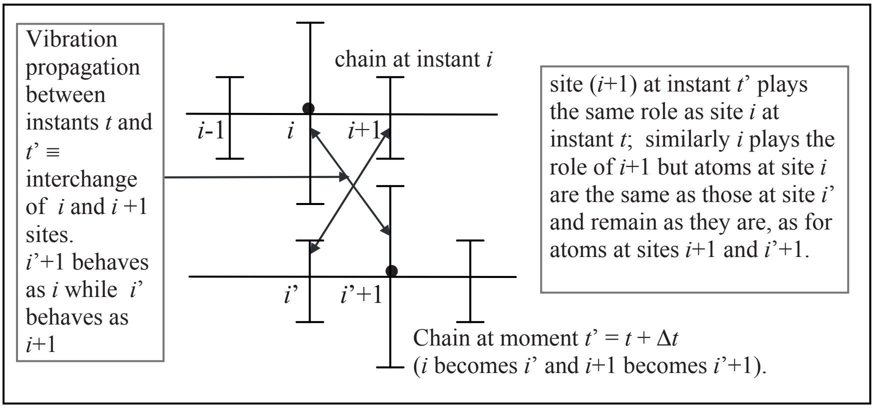
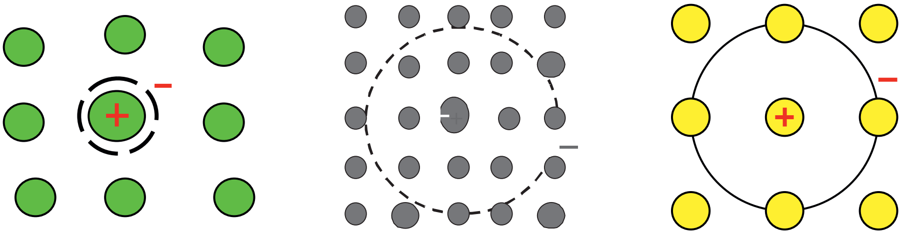


Figure 10.13. *"Permutation" of two neighboring sites on propagation and vibration (and with the polaron!) where i denotes the initial state and i' the final*

10.4. Excitons

10.4.1. *Physical origin*

In simple terms, just as an electron and a positively charged default are tied by an electrostatic force in a solid, an electron and a hole generated by an excitation can be attracted to form a tied state – or quasi-particle – termed an exciton. Depending on the nature of the solid, the excited state of electron and hole pair can be localized on one or more molecules. The former is called a Frenkel exciton, and is detailed in Figure 10.14a. For an electron and hole separated over several molecules, the result is called a Wannier exciton (Figure 10.14b). The intermediate between these two is the charge transfer exciton, where electron and hole are on adjacent molecules, as shown in Figure 10.14c.



(a) (b) (c)

Figure 10.14. *Various excitons: (a) Frenkel; (b) Wannier; and (c) charge transfer exciton*

10.4.2. *Wannier and charge transfer excitons*

In solids such as semiconductors that have large permitted bands (*B*), electrons and holes exhibit high mobilities (μ) with μ being proportional to *B* (as detailed in band theory for covalent solids). Holes and electrons can easily separate, meaning low interaction energies.

Given that an electron and hole pair, with energy levels schematically illustrated in Figure 10.16a are:

– buried within a continuous medium with permittivity given as $\varepsilon = \varepsilon_0\varepsilon_r$ where ε_r is relatively high for a reasonably well conducting medium (quite high mobility); and

– trace an orbit around one another in the material; thus this pair can be compared to the hydrogen atom where the nucleus is represented by the hole. The exciton is generally termed a Wannier exciton. With the energy levels quantified, and located with respect to the width of the gap of the material (E_G) (see Figure 10.15), they are given by:

$$E_{Gn} = E_G - \frac{m_{ex}^* e^4}{32\pi^2\varepsilon_0^2\varepsilon_r^2\hbar^2}\frac{1}{n^2}$$

where m_{ex}^* is the reduced effective mass of the electron-hole system and defined by the expression $m_{ex}^{*-1} = m_e^{*-1} + m_t^{*-1}$ where m_e^* and m_t^* are, respectively, the effective masses of the electron and the hole.

The dissociation of the exciton can be thought of as $n \to \infty$ and places the electron and hole into free states with the electron in the conduction band. Their interaction energy is equal to zero but they are separated by an energy given by $E_G = E_{G\infty}$. If the excitonic state corresponds to a given level denoted *n*, the electron and the hole are thus tied by a bonding energy E_{bn} such that

$$E_{bn} = E_G - E_{Gn} = \frac{m_{ex}^* e^4}{32\pi^2\varepsilon_0^2\varepsilon_r^2\hbar^2}\frac{1}{n^2}$$

In other words, this energy is what would be required to separate the electron and the hole from their bands where they are free (and such that $n \to \infty$, $E_{bn} \to 0$).

With the energy of separation for an electron–hole pair in an exciton being given as E_{Gn}, the lowest frequency absorption corresponds to the energy transition E_{G1} rather than E_{b1} (bonding energy).

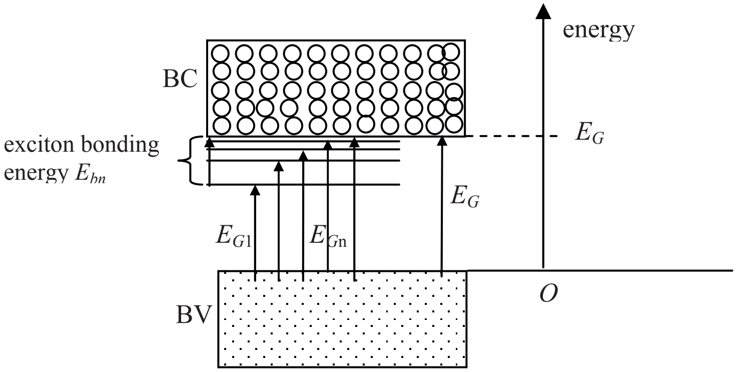


Figure 10.15. *The E_{Gn} energy level of an exciton with respect o the bonding energy level E_{bn}*

The radius of the exciton (r_n) can be evaluated using the electron–hole distance for a system based on hydrogen and is therefore written as $r_n = n^2\ (4\pi\varepsilon_0\varepsilon_r\hbar^2/\mathrm{m^*e^2})$. The distance r_1 ($n = 1$) is the shortest as we have the lowest energy level for a Wannier exciton and thus the most bonded exciton state (like an 1 *s*-orbital). From the expressions for E_n and r_n, we can see that bonding strength is essentially dependent on two parameters:

– the dielectric permittivity of the medium: the larger it is, the weaker the electron–hole attraction; and

– the reduced effective mass (m_{ex}^*) for electron-hole pairs: the smaller it is, the greater the electron–hole distance (and by consequence the harder it is to retain the exciton).

Thus, for semiconductors which exhibit a large permitted band, high permittivities and charge mobilities μ (with a low effective mass m^* as, classically, $\mu = q\tau/m^*$), only low bond energies appear (Wannier excitons). For semiconductors with an indirect gap, excitations of the lowest energy are forbidden as they require phonon intervention. For example, Wannier excitons have only been observed in semiconductors such as GaAs that have a direct gap.

Wannier excitons are theoretically possible for polymers. The bonding energy of the exciton is of the order of 0.4 eV and exhibits an ellipsoid geometry. These values are representative of molecular excitons with a high degree of localization on a chain under strong electron–lattice interactions (and electronic and vibrational state coupling). Polymers with a wide gap (insulators) must, nevertheless, present a weaker intra-chain delocalization and inter-chain interactions can be greater. In effect, charge transfer excitons (intermediate to Wannier and Frenkel excitons) are possible. The optimization of inter-chain contacts can result in excitons termed excimers (excitons shared over several identical molecular units) or exciplexes (excitons shared over two or more different molecular units).

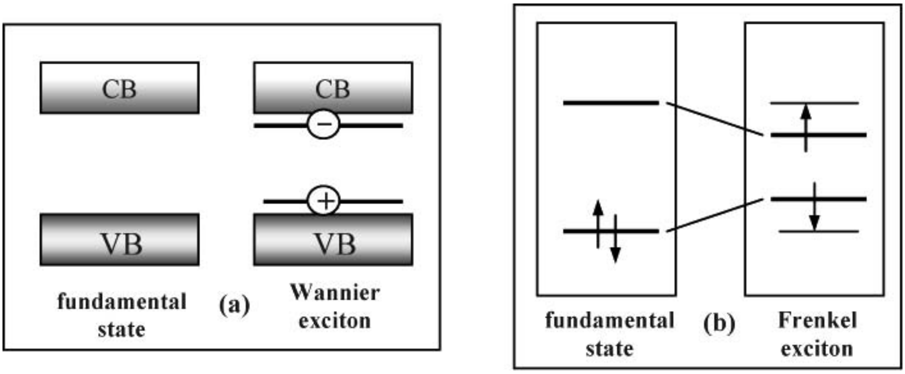


Figure 10.16. *Representations of excited states: (a) in a classic band scheme for semiconductors with a Wannier exciton; and (b) in a molecular state with discrete levels for a Frenkel exciton*

10.4.3. *Frenkel excitons*

In molecular crystals, the covalent bonds between atoms that make up the molecule are much stronger than intermolecular van der Waals bonds. Transitions between electronic levels of a practically isolated molecule in a dilute state and one in a condensed, solid state are only slightly changed in terms of frequency. Frenkel excitons are generally used to explain luminescence phenomena in molecular crystals through an excited state on a strongly bonded host molecule. The corresponding energy scheme is shown in Figure 10.16b.

Frenkel excitons are thus in a strong bond approximation with the excitation localized on the same molecule or on an adjacent neighbor. They have been observed during π–π^* transitions in aromatic molecules such as anthracene. Another example is that of excitations in ligand fields of d electrons, such as for nickel oxide.

While the particles that make up the electron-hole pair are bonded to one another on the same lattice site, together they constitute a quasi-particle which can move through the crystal by transferring energy to neighboring sites. This representation of energy migration is confirmed by the fact that in crystals which contain impurities, the excitations are trapped. It is in this manner that anthracene (which fluoresces on optical excitation), once doped with several parts per million of tetracene, sees its own fluorescence decrease and that of tetracene appear. This effect demonstrates that tetracene efficiently traps excitons, which can move across relatively large distances without relaxing. Given the relatively low concentrations of impurities, the distance covered can be of the order of a hundred molecules or more.

The movement of excitons can be due to overlapping of orbitals between adjacent sites. However, it should be noted that during the permitted π–π* transitions, it is the electrostatic interaction tied to the dipolar momentum which gives rise to the greatest coupling energy between fundamental and excited states. This term is expressed in the form $\mu_{ij}{}^2/R^3$ in which μ_{ij} is the dipolar momentum for the transition between fundamental (i) and excited (j) states.

For adjacent molecules of different types, the excitation energy transfer mechanisms with long (dipolar interactions) and short (orbital overlapping) action radii will be detailed, along with Förster and Dexter transfers. These transfers help explain the interest in the use of optical doping of fluorescent molecules especially for organic LEDs and can also help get round selection rules to increase LED yields.

10.5. Plasmons

10.5.1. *Basic definition*

A plasma is a medium in which the concentrations of positive and negative charges are equal and at least one of the charges is mobile. In a solid, the negative conduction charges (electrons) have a concentration equal to that of the positive ions.

10.5.2. *Dielectric response of an electronic gas: optical plasma*

10.5.2.1. *Dielectric function*

To a first approximation, a solid contains a fixed number n of ions per unit volume with n free electrons that are placed in a vacuum. A good example of such a material could be an alkaline metal. The dielectric response of the electron gas to an electric field that is such that $\vec{E} = \underline{E}_0 e^{j\omega t} \vec{e}_x = E_0\, \mathrm{e}^{j(\omega t - kz)}\, \vec{e}_x$ is obtained by working out the integration of the fundamental dynamic equation. We can assume that the electrons are not subject to recall or frictional forces. The equation, with respect to *Ox* is written as:

$$m\frac{d^2x}{dt^2} = -eE$$

From this equation, looking for forced solutions of the form $x = x_0 e^{j\omega t}$ gives us

$$x = \frac{eE}{m\omega^2}$$

The movement over a distance *x* by a charge (denoted -*e*) generates a moment dipole given by $p = -ex$, and the polarization (the dipole moment per unit volume) is therefore $P = np$. This means that

$$P = -\frac{ne^2}{m\omega^2}E$$

The dielectric function is such that $D() = \varepsilon_0 \varepsilon_r(\omega) E(\omega) = \varepsilon_0 E(\omega) + P(\omega),$ so

$$\varepsilon_r(\omega) = 1 + \frac{P(\omega)}{\varepsilon_0 E(\omega)} = 1 - \frac{ne^2}{\varepsilon_0 m\omega^2}$$

With the plasma frequency being defined by $\omega_p^2 = \frac{ne^2}{\varepsilon_0 m}$, the dielectric function can be written as:

$$\varepsilon_r(\omega) = 1 - \frac{\omega_p^2}{\omega^2}$$

10.5.2.2. *Optical plasma*

The equation, $\underline{k}^2 c^2 = \omega^2 \varepsilon_r(\omega)$, for the dispersion of electromagnetic waves shows that there are two differing regions (see Figure 10.17). If:

– $\omega > \omega_p$, $\varepsilon_r(\omega) > 0$ and the wave number (*k*) is real, then the wave $E = E_0\, e^{j(\omega t - kz)}$ is progressive and the dispersion is given by $\omega = \sqrt{\omega_p^2 + c^2 k^2}$.

– $\omega < \omega_p$, $\varepsilon_r(\omega) < 0$ and $\underline{k}$ is imaginary, as in $\underline{k} = -jk''$ then the wave $E = E_0\, e^{j(\omega t - \underline{k}z)} = E_0 e^{-k''z} e^{j\omega t}$ is evanescent, i.e. it has no propagation term but just terms for oscillation and attenuation at the surface.

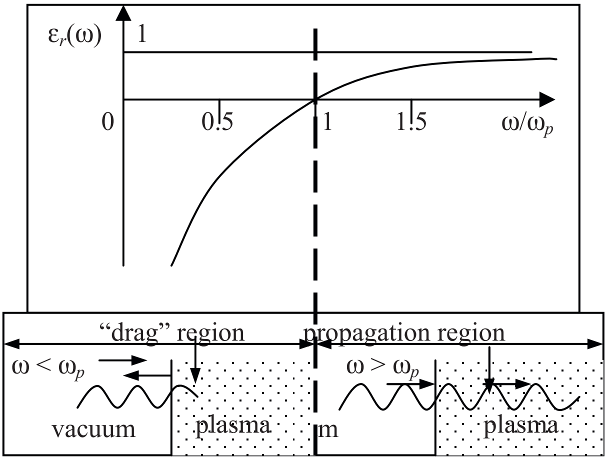


Figure 10.17. *High band filtering characteristics of a plasma*

In effect, the plasma behaves much as a high-band filter. As $\omega > \omega_p$ corresponds to $\lambda < \lambda_p$, waves of lengths less than λ_p can pass through the medium with $k'' = 0$, that is, without attenuation (without absorption). Alkaline metals where $\lambda_p \approx 300$ nm should therefore be transparent to wavelengths less than 300 nm, i.e. ultra-violet light.

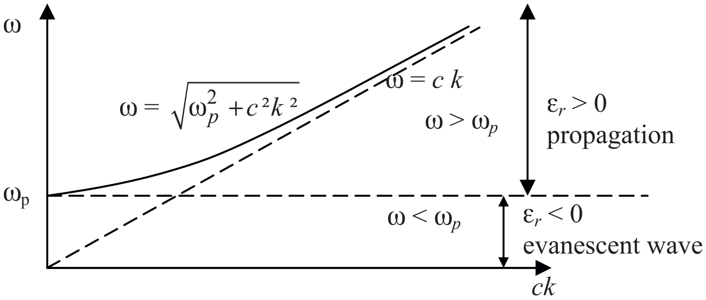


Figure 10.18. *Plasma dispersion curve*

For $\omega = \sqrt{\omega_p^2 + c^2k^2}$, the transverse wave dispersion curve is shown in Figure 10.18. The group velocity, $v_g = \frac{d\omega}{dk}$, is the dispersion curve slope and is less than the slope of the curve $\omega = ck$, and hence lower than c.

10.5.2.3. *Longitudinal optical modes in a plasma*

The function $\varepsilon(\omega)$ is equal to zero at a frequency denoted ω_L, so when

$$\left[\varepsilon_r(\omega)\right]_{\omega=\omega_L} = \left[1 - \frac{\omega_p^2}{\omega^2}\right]_{\omega=\omega_L} = 0\,,$$

then also $\omega_L = \omega_p$. A solution can arise for a longitudinal wave moving along x (with a wave vector $\vec{k}$). For more details see a course of electromagnetism, also section 2.6.2, problem 6.

Therefore there is a mode of longitudinal oscillation for a gas of electrons in a plasma. The angular frequency (or pulsation) ω_L of these oscillations is in fact equal to the plasma pulsation ω_p which in turn corresponds to the cutting frequency of the transverse electromagnetic waves.

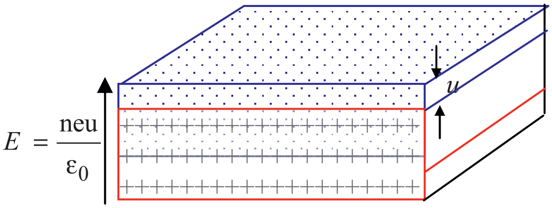


Figure 10.19. *Sections of a plasma placed in longitudinal oscillations with elongations denoted by u*

Physically speaking, we can expect that a longitudinal oscillation of a plasma through an elongation given by u will displace, for example, the electron gas upwards. The result of this negative charge in a section of width u is the formation of a layer with a surface charge given by $\sigma = -\text{neu}$ (the charge found in a parallelepiped with a unit surface and a height u tending to zero). Similarly, at the bottom there is a surface layer with a charge $\sigma = +\text{neu}$. A field $E = \frac{\text{neu}}{\varepsilon_0}$ is generated inside the layer, which tends to push the electrons back to their equilibrium position.

The field makes a recall force which is such that the dynamic fundamental equation relative to a unit volume of plasma can be written as:

$$nm\frac{d^2u}{dt^2} = -neE = -\frac{n^2e^2u}{\varepsilon_0}$$

so that with $\omega_p^2 = \dfrac{ne^2}{\varepsilon_0 m}$ we have:

$$\frac{d^2u}{dt^2} + \omega_p^2 u = 0.$$

This is the equation for a harmonic oscillator with an angular frequency ω_p, otherwise called the plasma frequency. The same value of ω_p is found via an alternate route in section 10.5.2.1.

10.5.3. *Plasmons*

10.5.3.1. *Definition and generation*

From what we have just seen, a plasma oscillation is a collective longitudinal oscillation of a conducting electron gas. By definition, the quantum energy $\hbar\omega_p$ of this collective plasma oscillation is called a plasmon, and $\omega_p = \left(\dfrac{ne^2}{\varepsilon_0 m}\right)^{1/2}$.

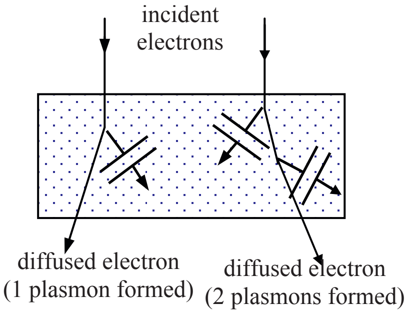


Figure 10.20. *Generation of plasmons by electron diffusion*

The pulsation, ω_p, is of the order of 10^{16} rad s^{-1}, and the energy of a plasmon in a wide range of solids can vary from around 3 to 30 eV. Simple thermal excitation of plasma oscillations is therefore impossible ($kT \approx 0.026$ eV for $T = 300$ K). To excite plasmons it is necessary to, for example, have a flux of electrons penetrate the lattice. On moving to the interior of the crystal, each electron can excite several successive plasmons. Each plasmon is symbolized by —| |→.

10.5.3.2. *Some properties*

The high energies of plasmons can be explained by their oscillations being the result of a high number of electrons. The excitation of plasmons can be obtained either by making electrons (or ions or photons) traverse or reflect the film under study.

In a dielectric, it is also possible to excite collective plasma oscillations. Here it is the valence electron cloud that as a whole oscillates with respect to the positive ions.

In semiconductors, two plasma oscillation modes can arise. One is at high frequencies and involves all the valence electrons in an oscillation. The other, a low frequency oscillation, is associated with the electrons in the conduction band where the plasma pulsation given by $\omega_p{}' = \left(\frac{ne^2}{\varepsilon_0 \varepsilon_r m^*}\right)^{1/2}$ is such that $\hbar\omega_p'$ (≈ 0.01 eV) is small.

At the interface between a metal and a dielectric (for example in the case of a metal oxide layer on a metal) the surface plasmons can be excited and propagate with an energy that is less than those in the bulk. The study of surface plasmons makes it possible to gain information on the permittivity and thickness of the surface insulating layer.

10.6. Problems

10.6.1. *Problem 1: enumeration of vibration modes (phonon modes)*

This problem is based on a volume V of sides L_x, L_y and L_z belonging to a simple cubic system. It encloses N cubic elementary cells of sides a:

1) Determine the number of the vibrational modes.

2) Calculate the density of the vibration modes for each type of polarization by using the Debye approximation where $\omega = v_s|K|$.

Answers

1) For the cubic crystal (of sides L_x, L_y, L_z), the periodic limiting conditions applied to the three components of the vector $\vec{K}$ gives $e^{i\left[K_x x + K_y y + K_z z\right]} = e^{i\left[K_x(x+L)+K_y(y+L)+K_z(z+L)\right]}$ and hence the values $K_x = \frac{2\pi}{L_x} n_x$, $K_y = \frac{2\pi}{L_y} n_y$, $K_z = \frac{2\pi}{L_z} n_z$. The extremities of the $\vec{K}$ wave vectors are found at the nodes of an orthorhombic lattice based on $2\pi/L_x$, $2\pi/L_y$, $2\pi/L_z$. There are three nodes associated with each node ($1L + 2T$), and these three nodes give the elemental cell with a volume given by $\frac{2\pi}{L_x}\frac{2\pi}{L_y}\frac{2\pi}{L_z} = \frac{8\pi^3}{V}$. The number ($N_{3m}$) of these cells (associated with three nodes) that can be placed in the first Brillouin zone with a volume of $\left(\frac{2\pi}{a}\right)^3$ in a cubic lattice is thus given by $\left(\frac{2\pi}{a}\right)^3 \Big/ \frac{8\pi^3}{V} = \frac{V}{a^3} = N$, where N is the number of cells in the direct lattice, and hence also the number of atoms in the lattice (if on average there is one atom per crystal cubic cell). Definitely, we have $N_{3m} = N$, and as N_{3m} is the number of cell associated with three modes, then the total number of modes is $3N$.

2) For a given polarization, we have a mode (with a given value of $|K|$) associated with a cell of volume $\frac{2\pi}{L_x}\frac{2\pi}{L_y}\frac{2\pi}{L_z} = \frac{8\pi^3}{V}$. with a unit of reciprocal space, there is therefore associated $\frac{1}{8\pi^3/V} = \frac{V}{8\pi^3}$ modes/polarization.

In addition, the mode density $D(K)$ (where each is associated with a K value) must be such that $D(K)dK$ represents the number of modes characterized by a value of K between K and $K+dK$. These K values must therefore be found between the spheres of radius K and $K + dK$, i.e. within a volume $4\pi K^2 dK$.

In this volume can be placed $\frac{4\pi K^2 dK}{8\pi^3/V} \overset{V=1}{=} \frac{K^2 dK}{2\pi^2}$ cells for each mode (for a certain polarization). We can set $V = 1$ so as to calculate the density. From this we find $D(K) = \frac{K^2}{2\pi^2}$.

As we should have $D(\omega)d\omega = D(K)dK$, we deduce that $D(\omega) = D(K)\frac{dK}{d\omega}$, so that with the Debye approximation where $\omega = v_s|K|$ (and $\frac{dK}{d\omega} = \frac{1}{v_S}$) we find $D(\omega) = \frac{\omega^2}{2\pi^2 v_S^3}$ for the mode density for each type of polarization.

10.6.2. *Problem 2: polaritons*

In a 3D crystal there is a chain of ions separated by a distance denoted a. There are two types of ions; the positive and negative charges are alternatively placed. Ions with a charge $+e$ and mass m are situated at $z = 2na$, while ions with the charge $-e$ and mass M are placed at $z = (2n \pm 1)a$. Make a diagram of this.

The system is subjected to a transverse sinusoidal electric field $\vec{E}_x \parallel \overrightarrow{Ox}$ which propagates along Oz. The equation for its form is $E_x = E_0 \exp\left[j\left(\omega t - kz\right)\right]$, where the angular frequency is in the infra-red (IR) range, i.e. $\omega \approx 10^{13}$ rd s^{-1}.

1) Show that the IR wavelength is very large with respect to a ($\approx$ 0.3 nm), and this makes it possible to neglect the propagation term in the expression for the IR wave. Write the new, approximate form of the equation, which will then be used in the following questions.

2) Taking only interactions between nearest neighboring ions, give the expression for the spread (A–B) in the distances between the two types of ions A and B. The result should be expressed as a function of the masses of the ions, their charges and angular frequencies (ω) and $\omega_T = \sqrt{\frac{2\beta}{\mu}}$, where β is the coupling constant between ions and μ is their reduced mass.

3) Give the general equation for the dipole moment with respect to x for a system based on two ions and their nearest neighbors. From this deduce the model for the ionic polarization vector in a solid containing N pairs of ions (of the type studied) per unit volume.

4)

a) Establish the expressions for the dielectric permittivity [ε (ω)] as a function of N, μ, ω, ω_T, and ε_∞.

b) Introduce the parameter $\Omega_p^2 = \frac{Ne^2}{\mu}$ into the equation for $\varepsilon(\omega)$ and from this deduce the (Born) equation for the static permittivity [$\varepsilon(0)$] which is such that $\varepsilon(0) = \varepsilon_s$ by notation. Finally, express $\varepsilon(\omega)$ as a function of ε_s, ε_∞, ω_T, and ω.

5) Make $\omega_L^2 = \omega_T^2 \frac{\varepsilon_S}{\varepsilon_\infty}$ (in the Lyddane-Sachs-Teller equation). Express ε(ω) as a function of ε_∞, ω_T, ω_L, and ω. From it deduce the physical significance of ω_L, and then the domain of the angular frequency (otherwise called pulsation) for which the electric waves are reflected by the crystal.

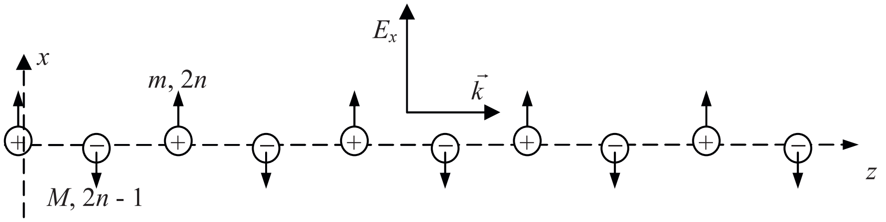
6) Plot the curve ε(ω) = *f*(ω). Study, in particular, the pole and the zero points of the dielectric function ε(ω). Show how the pulsation (ω_L) corresponds to a longitudinal mode wave. Detail the consequence of this.

7) Show that in an ionic crystal where both phonons (vibrational movements) and plane transverse electromagnetic waves in the IR region that the dispersion equation can be written as $\varepsilon_{r\infty}\omega^4 - \omega^2\left(\varepsilon_{rs}\omega_T^2 + k^2c^2\right) + k^2c^2\omega_T^2 = 0$. Note that the IR wave is in the domain of the proper pulsation ($\omega_T = \sqrt{\frac{2\beta}{\mu}}$) of transverse optical phonons, and the index *r* of the permittivities designates that they are relative dielectric permittivities.

8) The quantum of fields coupled with photons (transverse electromagnetic waves, $E_x = E_0 \exp\left[j\left(\omega t - kz\right)\right]$) and of transverse phonons is called a polariton. Show graphically the dispersion curves for these fields that are initially uncoupled and then coupled.

Answers

1)



We have $\lambda = \frac{c}{\nu} = \frac{2\pi c}{\omega}$, so with ω ≈ 10^{13} rd s^{-1}, we find that λ ≈ 200 µm. With *a* ≈ 0.3 nm it is possible to state that λ >> *a*. The wave remains practically constant over a large part of the *z* length of the chain – and makes it possible to neglect the term for the propagation. The form of the wave can thus be reduced, by neglecting the propagation term, to $E_x = E_0 \exp j\omega t$. Mathematically speaking, we can also say that as $k = 2\pi/\lambda$ is very small especially when λ is becomes very large with respect to z, the product kz tends towards zero, and therefore the term $\exp(-jkz) \to \exp 0 \equiv 1$ and E_x is reduced to:

$$E_x = E_0 \exp j\omega t \exp(-jkz) \to E_0 \exp j\omega t$$

2) The movement equations are forced by the electric field to give:

– for atoms of mass m, we have

$$m\frac{d^2u_{2n}}{dt^2} = \beta\left(u_{2n+1} + u_{2n-1} - 2u_{2n}\right) + eE_0e^{j\omega t}$$

– and for atoms of mass M, then

$$M\frac{d^2u_{2n+1}}{dt^2} = \beta\left(u_{2n+2} + u_{2n} - 2u_{2n+1}\right) - eEe^{j\omega t}$$

The search for forced solutions written as $u_{2n} = Ae^{j\omega t}$ and $u_{2n+1} = Be^{j\omega t}$, substituted in to the preceding equations gives:

– $m\omega^2 A = 2\beta\left[B - A\right] + eE_0$

– $M\omega^2 B = 2\beta\left[A - B\right] - eE_0$

By dividing the first equation by m and the second by M, then subtracting term by term, we find that by making $\frac{1}{\mu} = \frac{1}{m} + \frac{1}{M}$ and $\omega_T^2 = \frac{2\beta}{\mu}$ (frequency of the branch of optical phonons when $K \to 0$) that:

$$A - B = \frac{eE_0}{\mu\left[\omega_T^2 - \omega^2\right]} \qquad [10.20]$$

The equation that is thus obtained shows that there is a resonance frequency attained when $\omega = \omega_T$ (between the excitation of the pulsation wave ω and the optical phonons when $K \approx 0$).

3) The negative and positive charged ions are therefore displaced in the opposite sense to the electric field and generate a dipole moment of an amplitude $|m_e| = e(A - B)$. By definition, the polarization is the dipole moment per unit volume. If N represents the number of ion pairs (and hence dipoles) per unit volume, then the modulus of the ionic polarizability (associated with the displacement of only the ions) is $P_{\text{ion}} = N\, m_e\, (A - B)$, so that:

$$P_{\text{ion}} = \frac{Ne^2}{\mu\left[\omega_T^2 - \omega^2\right]} E_0 \qquad [10.21]$$

4)

a) In general terms, $P = \left(\varepsilon - \varepsilon_0\right) E$. The polarization $P(\omega)$ of a given pulsation ω can be composed of two terms:

– one for the polarization which establishes itself instantaneously and is associated with the electronic polarization (where the electrons can instantaneously follow and applied field) and takes on a frequency $\nu = [1/(t \to 0)] = \infty$: $P_\infty = \left(\varepsilon_\infty - \varepsilon_0\right) E$;

– and another (P_{ion}) here associated with a slower ionic polarization that that above.

We therefore have:

$$\begin{aligned} P(\omega) &= \left[\varepsilon(\omega) - \varepsilon_0\right] E_0 \\ &= P_\infty + P_{\text{ion}} = \left[\varepsilon_\infty - \varepsilon_0\right] E_0 + \frac{Ne^2}{\mu\left[\omega_T^2 - \omega^2\right]} E_0 \end{aligned}$$

b) With $\Omega_p^2 = \frac{Ne^2}{\mu}$, and $\omega_p^2 = \frac{Ne^2}{\varepsilon_0 \mu}$ where ω_p is the plasma pulsation, we have:

$$\varepsilon(\omega) = \varepsilon_\infty + \frac{Ne^2}{\mu\left[\omega_T^2 - \omega^2\right]} = \varepsilon_\infty + \frac{\Omega_p^2}{\left[\omega_T^2 - \omega^2\right]} \qquad [10.22]$$

From equation [10.22] it is possible to deduce the Born equation, as in:

$$\varepsilon(0) \overset{\text{notation}}{=} \varepsilon_S = \varepsilon_\infty + \frac{\Omega_p^2}{\omega_T^2} \qquad [10.23]$$

The result is $\Omega_p^2 = [\varepsilon_S - \varepsilon_\infty]\omega_T^2$. Taking the last equation and placing it into equation [10.22], we have:

$$\varepsilon(\omega) = \varepsilon_\infty + \frac{\omega_T^2}{\left[\omega_T^2 - \omega^2\right]}[\varepsilon_S - \varepsilon_\infty] \qquad [10.24]$$

5) Here we make:

$$\omega_L^2 = \omega_T^2 \frac{\varepsilon_S}{\varepsilon_\infty} \qquad [10.25]$$

In passing, we can write equation [10.25] in another form, i.e.:

$$\frac{\varepsilon_\infty}{\varepsilon_S} = \frac{\omega_T^2}{\omega_L^2} \qquad [10.26]$$

and it is this that makes up the so-called Lyddane-Sachs-Teller equation and in which we can determine the physical significance of ω_L.

To make equation [10.25] appear in [10.24] we can write:

$$\begin{aligned}\varepsilon(\omega) &= \varepsilon_\infty \left[1 - \frac{\omega_T^2}{\left[\omega_T^2 - \omega^2\right]}\right] + \varepsilon_S \frac{\omega_T^2}{\omega_T^2 - \omega^2} \\ &= \varepsilon_\infty \frac{\omega_T^2 - \omega^2 - \omega_T^2 + \omega_T^2\left(\varepsilon_S / \varepsilon_\infty\right)}{\omega_T^2 - \omega^2}\end{aligned}$$

So that with equation [10.25]:

$$\varepsilon(\omega) = \varepsilon_\infty \frac{\omega_L^2 - \omega^2}{\omega_T^2 - \omega^2} \qquad [10.27]$$

And then we see straight away that $\left[\varepsilon(\omega)\right]_{\omega=\omega_L} = 0$.

As $\varepsilon_s > \varepsilon_\infty$, then according to equation [10.25] $\omega_L > \omega_T$. The result is that when $\omega_T < \omega < \omega_L$, we have $\varepsilon(\omega) < 0$. In this case, the relation for the dispersion of transverse electromagnetic waves:

$$k^2 = \frac{\omega^2}{c^2}\varepsilon_r(\omega) = \frac{\omega^2}{c^2}\frac{\varepsilon(\omega)}{\varepsilon_0},$$

shows that when $\varepsilon(\omega) < 0$, $k^2 < 0$, then $k = -ik''$ must be true and contains a pure imaginary number. The wave thus becomes:

$$E_x = E_0 \exp[-k''r]\exp[j\omega t]$$

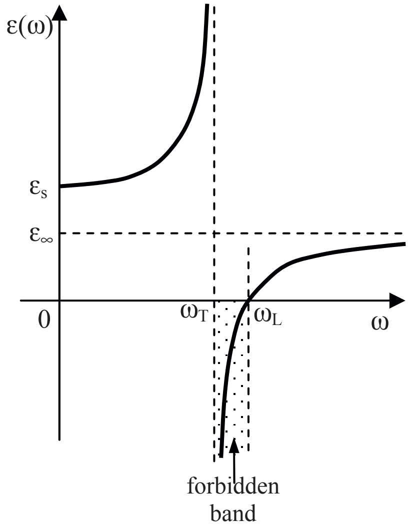
and is retarded (evanescent) in the material and no longer propagates. This corresponds to a forbidden band for pulsation waves between ω_T and ω_L.

6) Especially when using equation [10.27] we can see that the limiting values are as follows:

- when $\omega = 0$, we have $\varepsilon(\omega) = \varepsilon_s$;
- when $\omega \to \infty$, $\varepsilon(\omega) \to \varepsilon_\infty$;
- when $\omega \to \omega_T^-$, $\varepsilon(\omega) \to +\infty$;
- when $\omega \to \omega_T^+$, $\varepsilon(\omega) \to -\infty$;
- when $\omega = \omega_L$, $\varepsilon(\omega) = \varepsilon(\omega_L) = 0$.

From this it is possible to deduce the curve shown in the figure on the next page.

The highest point for the dielectric function is at ω_T (resonance frequency between the IR wave and the transverse optical phonons). The zero point for this dielectric function is obtained when $\omega = \omega_L$.



In general terms, the Gauss equation, $\text{div}\vec{D} = \varepsilon(\omega)\text{div}\vec{E} = 0$, has in fact two solutions:

– either $\text{div}\vec{E} = 0$ (and hence ε (ω) ≠ 0, which happens when $\omega \neq \omega_L$), from which we deduce that $i\vec{k}.\vec{E} = 0$. We therefore have $\vec{E} \perp \vec{k}$ which corresponds to a transverse wave structure;

– or ε(ω) = 0, which is the precise result when $\omega = \omega_L$. In this case, the Gauss equation is true without the wave necessarily being transverse (as $\text{div}\vec{E} = 0$ no longer has to be true). We can now go on to look for the form of the wave.

Here, as usual, we have:

$$\left.\begin{array}{l} \text{div}\vec{B} = 0 \Rightarrow i\vec{k}.\vec{B} = 0 \\ \text{rot}\vec{B} = \mu_0\varepsilon(\omega)\dfrac{\partial \vec{E}}{\partial t} = j\omega\mu_0\varepsilon(\omega)\vec{E} \overset{\omega=\omega_L}{=} 0 \Rightarrow i\vec{k}\times\vec{B} = 0 \end{array}\right\} \Rightarrow B = 0$$

The wave can therefore only be purely electrical. The Maxwell-Faraday equation makes it possible to state that $\overrightarrow{\text{rot}}\vec{E} + \frac{\partial \vec{B}}{\partial t} = 0$, so that in addition, $\overrightarrow{\text{rot}}\vec{E} + j\omega\vec{B} = 0$. With $\vec{B} = 0$, we have $\overrightarrow{\text{rot}}\vec{E} = 0$, and $j\vec{k}\times\vec{E} = 0$, and hence $\vec{E} \parallel \vec{k}$. This means that $\vec{E}$ is directed along $\vec{k}$ and the wave has a longitudinal structure. Finally the pulsation denoted ω_L appears either as the upper limit for the forbidden band, either as the pulsation with which is associated a longitudinal phonon.

7) Taking the equation for the wave dispersion, $k^2 = \frac{\omega^2}{c^2}\varepsilon_r(\omega) = \frac{\omega^2}{c^2}\frac{\varepsilon(\omega)}{\varepsilon_0}$, the expression given in [10.27] can be written dividing through the members with ε_0, as in $\varepsilon_r(\omega) = \varepsilon_{r\infty}\frac{\omega_L^2-\omega^2}{\omega_T^2-\omega^2}$. We obtain: $k^2c^2\omega_T^2 - k^2c^2\omega^2 = \varepsilon_{r\infty}\left(\omega_L^2-\omega^2\right)\omega^2$, such that $\varepsilon_{r\infty}\omega^4 - \omega^2\left(\varepsilon_{r\infty}\omega_L^2 + k^2c^2\right) + k^2c^2\omega_T^2 = 0$. Using the Lyddane-Sachs-Teller equation, i.e. $\varepsilon_{r\infty}\omega_L^2 = \varepsilon_{rs}\omega_T^2$, we find:

$$\varepsilon_{r\infty}\omega^4 - \omega^2\left(\varepsilon_{rs}\omega_T^2 + k^2c^2\right) + k^2c^2\omega_T^2 = 0 \qquad [10.28]$$

8) The solutions for equation [10.28] are in the form:

$$\omega_\pm^2 = \frac{1}{2\varepsilon_{r\infty}}\left[\left(\varepsilon_{rs}\omega_T^2 + k^2c^2\right) \pm \sqrt{\left(\varepsilon_{rs}\omega_T^2 + k^2c^2\right)^2 - 4k^2c^2\varepsilon_{r\infty}\omega_T^2}\right] \qquad [10.29]$$

In addition, the dispersion relation $k^2 = \frac{\omega^2}{c^2}\varepsilon_r(\omega)$ associated with equation [10.27] written in the form $\varepsilon_r(\omega) = \varepsilon_{r\infty}\frac{\omega_L^2-\omega^2}{\omega_T^2-\omega^2}$, shows that:

– when k ≈ 0, there are two possible solutions:

- either ω ≈ 0, which corresponds to the smallest obligatory solution for equation [10.29], as can also be directly verified by placing $k \approx 0$ in to the equation for ω_-,

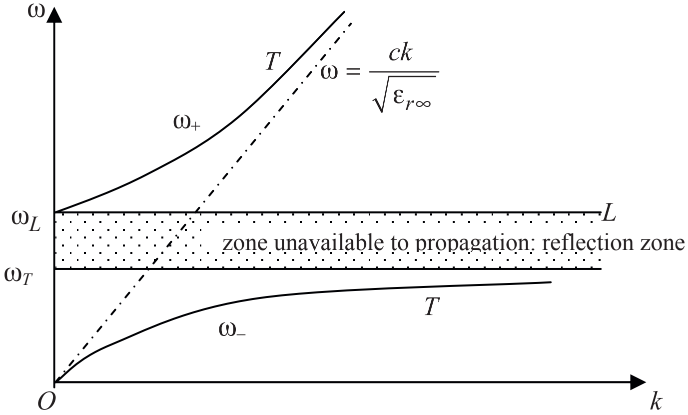
- $\varepsilon_r(\omega) \approx 0$, which can be obtained when $\omega \approx \omega_L$. This is also in accordance with the LST equation in which $\omega_L^2 = \frac{\varepsilon_S}{\varepsilon_\infty}\omega_T^2$. This solution corresponds to the solution ω_+ as can be directly verified in equation [10.29] when placing $k \approx 0$ in ω_+;

– when $k \to \infty$, there are two further solutions:

- either $\frac{\omega}{c} \to \infty$ and $k \approx \frac{\omega}{c}\sqrt{\varepsilon_{r\infty}}$ and hence the solution for ω_+. The introduction into equation [10.29] for $k \to \infty$, such that $k^2c^2 >> \varepsilon_{rs}\omega_T^2$, where $k^2c^2 >> \varepsilon_{r\infty}\omega_T^2$ and goes towards a solution for ω_+ as $\omega_+ = \frac{ck}{\sqrt{\varepsilon_{r\infty}}}$,

- or $\varepsilon_r(\omega) = \varepsilon_{r\infty} \frac{\omega_L^2 - \omega^2}{\omega_T^2 - \omega^2} \to \infty$, which is obtained when ω → ω_T, and corresponds for the solution for ω_- in equation [10.29].

The ω = *f*(*k*) plot shows two sets of curves associated with the transverse waves (set *T*) for two solutions ω_+ and ω−. These are the coupled modes for photons and transverse optical photons in an ionic crystal containing the polaritons.



As shown in the figure above, these two solutions are separated by a forbidden zone (or band) with respect to the propagation of the electromagnetic waves:

– the value ω = ω_L corresponds to a longitudinal wave (*L*) and is the vibration of longitudinal optical phonons;

– the value ω = ω_T is for a transverse wave (*T*) and is the vibration of transverse optical phonons.

The dotted line represents the (straight) dispersion curve for single photons in a crystal (uncoupled with the lattice vibrations).

This coupling effect, that modifies the resonance frequencies (for photons with angular frequencies ω with phonons with angular frequencies denoted ω_T and ω_L), the propagation property of the electromagnetic wave is therefore reflected when ω is such that $\omega_T < \omega < \omega_L$.

In this zone, the electromagnetic wave vector is purely complex. This can be considered as a breaking of the wave in the neighborhood of the surface of the solid.

Bibliography

[AND 58] ANDERSON P.W., "Absence of Diffusion in Certain Random Lattices", *Phys. Rev.*, 109, 5, p.1492–1505, 1958.

[ASH 76] ASHCROFT N.W and MERMIN D., *Solid State Physics*, Saunders College Publishing, 1976.

[BRO 92] BROUSSEAU M., *Physique du solide*, Masson, Paris, 1992.

[CAZ 96] CAZAUX J., *Initiation à la physique du solide*, Masson, Paris, 1996.

[FRI 70] FRITZSCHE H., "CFO Model", *la Recherche*, Novermber, p. 527, 1970.

[GER 97] GERL M. and ISSI J.P., *Physique des matériaux*, Presses Polytechniques et Universitaires Romandes, Lausanne, 1997.

[IOF 60] IOFFE A.F., REGEL A.R. "Non crystalline, amorphous and liquid electronic semiconductors", *Progress in Semiconductors*, vol. 4, Heywood and Co Ltd, London, 1960.

[KIT 63] KITTEL C. *Quantum Theory of Solids*, Wiley, New York, 1963.

[KIT 96] KITTEL C. *Introduction to Solid State Physics*, Wiley, New York, 1996.

[LEV 95] LEVY F., *Physique et technologie des semi-conducteurs*, Presses Polytechniques et Universitaires Romandes, Lausanne, 1995.

[MIL 60] MILLER A. and ABRAHAMS E., "Impurity Conduction at Low Concentrations", *Phys. Rev.*, 120, 3, p.754–755, 1960.

[MOL 06] MOLITON, A. *Optoelectronics Of Molecules And Polymers*, Springer, New York, 2006.

[MOL 07a] MOLITON A., *Applied Electromagnetism and Materials,* Springer, 2007.

[MOL 07b] MOLITON, A. *Basic Electromagnetism and Materials*, Springer, New York, 2007.

[MOO 93] MOOSER E., *Introduction à la physique des solides*, Presses Polytechniques et Universitaires Romandes, Lausanne, 1993.

[MOT 79] MOTT N.F. and DAVIS E.A., *Electronic Processes in Non-crystalline Materials,* Clarendon, 1979.

[QUE 88] QUERE Y., *Physique des matériaux*, Ellipses, Paris, 1988.

[SAP 90] SAPOVAL B. and HERMANN C., *Physique des semi-conducteurs*, Ellipses, Paris, 1990.

[SMI 67] SMITH R.A., *Wave Mechanics of Crystalline Solids*, Chapman & Hall, London, 1967.

[TEY 92] TEYSSIER J.L. and BRUNET H., *Physique des matériaux conducteurs et semi-conducteurs*, Dunod, Paris, 1992.

[VAP 70] VAPAILLE A., *Physique des dispositifs à semi-conducteurs, tome 1 : électronique du silicium homogène*, Masson, Paris, 1970.

Index